

Working Paper 93-32
Statistics and Econometrics Series 23
November 1993

Departamento de Estadística y Econometría
Universidad Carlos III de Madrid
Calle Madrid, 126
28903 Getafe (Spain)
Fax (341) 624-9849

MODELLING DAILY SERIES OF ECONOMIC ACTIVITY

Antoni Espasa*

Abstract

The behavior of daily series of economic activity like the consumption of electric energy, cash withdrawn from financial institutions, number of passengers in a transport service, pollution and traffic levels, sales, etc., is very often characterised by showing oscillating levels or trends and several complex seasonalities. Besides, these series are sensible to: (1) the presence of holidays; (2) vacation periods and; (3) the end and beginning of month. Finally, these series suffer, in general, from the influence of meteorological variables and in many cases the effects are nonlinear, dynamic and change with the type of the day -weekdays or weekends or holidays- and season of the year.

The levels of these series show so complex trends and oscillations that the process of their modelling becomes very difficult. Nevertheless, because such oscillations correspond to behaviour patterns of the economic agents the modelling task is not only feasible but also very rewarding.

The paper specifies the main characteristics of daily series of economic activity; analyzes how these features can be explained by a quantitative model; proposes a strategy for the construction of those models; illustrates their use for forecasting and control purposes and shows examples of models already in active use for several years that have up to almost two hundred estimated parameters employing several thousands observations.

Key Words

Forecasting; Seasonality; Calendar Effects; Meteorological Variables.

*Departamento de Estadística y Econometría, Universidad Carlos III de Madrid.

MODELLING DAILY SERIES OF ECONOMIC ACTIVITY

Antoni Espasa¹

Universidad Carlos III, Madrid, Spain

October, 1993

¹ The author wants to acknowledge financial support from the Spanish DGICYT contract PB-90-0267 and from the Argentaria Chair at the Universidad Carlos III de Madrid.

Abstract

The behavior of daily series of economic activity like the consumption of electric energy, cash withdrawn from financial institutions, number of passengers in a transport service, pollution and traffic levels, sales, etc., is very often characterised by showing oscillating levels or trends and several complex seasonalities. Besides, these series are sensible to: (1) the presence of holidays; (2) vacation periods and; (3) the end and beginning of month. Finally, these series suffer, in general, from the influence of meteorological variables and in many cases the effects are nonlinear, dynamic and change with the type of the day -weekdays or weekends or holidays- and season of the year.

The levels of these series show so complex trends and oscillations that the process of their modelling becomes very difficult. Nevertheless, because such oscillations correspond to behaviour patterns of the economic agents the modelling task is not only feasible but also very rewarding.

The paper specifies the main characteristics of daily series of economic activity; analyzes how these features can be explained by a quantitative model; proposes a strategy for the construction of those models; illustrates their use for forecasting and control purposes and shows examples of models already in active use for several years that have up to almost two hundred estimated parameters employing several thousands observations.

Key words: Forecasting, seasonality, calendar effects, meteorological variables.

INDEX

- I. Introduction.
- II. Characteristics of the daily series of economic activity and some first considerations for their modelling.
- III. Some examples of daily series.
- IV. The use of variance reduction criterium in daily modelling.
- V. Modelling the trend and the seasonals without indicators.
 - V.1. Modelling local oscillations and trends.
 - V.2. Modelling seasonalities by deterministic schemes.
 - V.3. Stochastic modelling of the seasonals.
- VI. Meteorological variables as indicators for annual seasonality.
- VII. End and beginning-of-month effect.
- VIII. Some examples in modelling trend and seasonalities in daily series.
- IX. Modelling calendar effects.
 - IX.1. Holidays
 - IX.2. Vacation periods
 - IX.3. Special events
- X. Dynamic and nonlinear effects of the meteorological variables.
- XI. The use of daily models for forecasting and control. Some practical examples for models in daily use for several years.

References

I. Introduction

The behavior of daily series of economic activity like the consumption of electric energy, water, gas, petrol, etc., cash withdrawn from financial institutions, bank deposits, eligible liabilities, notes and coin in circulation, number of passengers in a transport service, pollution and traffic levels, sales, etc., is very often characterised by showing oscillating levels or trends and several complex seasonalities of weekly, monthly, quarterly and annual periodicities. Besides, these series are sensible to: (1) the presence of holidays (specially important when they occur in the middle of the week) with influence depending on the day of the week, season of the year and number of people affected; (2) vacation periods like Easter, summer -month of August in several European countries- and Christmas; (3) the end and beginning of month, very often depending differently according to the month, season or presence of holidays; etc. Finally, these series suffer, in general, from the influence of meteorological variables and in many cases the effects are nonlinear, dynamic and change with the type of the day -weekdays or weekends or holidays- and season of the year.

The levels of these series show so complex trends and oscillations that the process of their modelling becomes very difficult. Nevertheless, because such oscillations correspond to behaviour patterns of the economic agents the modelling task is not only feasible but also very rewarding. This implies, as should be pretty obvious, a substantial difference with another type of daily data, those corresponding to yield series. These series can be considered as prices in efficient or quasiefficient markets and therefore the changes in their levels do not show systematic patterns in their evolution.

For yields and similar financial series the interesting quantitative problem is not the conditional modelling for the levels, but for the second moments. Quite to the contrary, for what it has been called at the beginning of this paper series of economic activity, the changes in levels are forecastable and an adequate model to explain them is very demanding and certainly gives definite advantage or benefit to the agent which possesses such a model. In fact, for the experts in firms or institutions with daily problems in forecasting economic activity data it is almost impossible to assimilate all the particularities that these data show. For

instance, the precise evaluation of the end of March effect as different from other months in a different season and as a function of the day of the week in which March ends, of the proximity of Easter and of the divergence of the values of different meteorological variables (temperature, humidity, sunshine, wind, etc.) from their normal values at this time of the year, is something that usually escapes from the control of the expert. But all these things can be included in a model which could be used to produce better forecasts than those from experts. In fact, the model based forecast will be very much better than a subjective forecast in those days affected by a combination of different calendar factors and with meteorological conditions quite away from normal. In many cases a good forecast for those days justifies the inversion on modelling.

The models are also important in firms and institutions with an expert which is able to produce good forecasts even in the more difficult days. By certain, the models could incorporate the knowledge of the expert and convert the forecasting task as a rutinary job. That saves time from qualified experts, avoid an excessive dependence of the firm from them, allows the access of other persons to the model and the knowledge which it incorporates and makes easier to learn from errors to producing better forecasts.

The aim of this paper is: (a) to specify the main characteristics of daily series of economic activity; (b) to analyze how these features can be explained by a quantitative model; (c) to propose a strategy for the construction of those models; (d) to illustrate their use for forecasting and control purposes and (e) to show examples of models already in active use for several years that have up to almost two hundred estimated parameters employing several thousands observations.

II. Characteristics of the daily series of economic activity and some first considerations for their modelling

The daily series of economic activity usually show several of the following characteristics:

(1) TRENDS

- a trend or locally oscillating level.

(2) SEASONAL OSCILLATIONS

- (a) weekly seasonality, (b) annual seasonality, (c) end and beginning-of-month effects,

(3) CALENDAR EFFECTS

- oscillations due to the presence of holidays, and changes in the trend and in the seasonals due to vacation periods.

(4) DEPENDENCE FROM EXOGENOUS VARIABLES

- complex dependence from exogenous explanatory variables, like meteorological variables.

In general these daily series show more than one of the above characteristics and sometimes the presence of one conceals others. So, it seems convenient to face the problem of modelling assuming that several of the pointed characteristics, or even another ones which usually appear less frequently, like quarterly seasonality, are present in the data and establishing a process to test if they are there or not. Thus, the starting point should be a listing of possible characteristics of the data and the following steps would consist on testing the presence of each one of them and in the affirmative case to model it.

The modelling process can be organized according with the following principles:

- (a) determination of orthogonalities or quasi-orthogonalities between the list of characteristics;
- (b) classification of these characteristics from the most influential in the data to the less with the aim of establishing an order in their modelling;
- (c) determination of the characteristics for which there are indicators and

- (d) selection of stochastic or deterministic schemes to model the characteristics for which an indicator is not available.

The principle (a) indicates that in a first step the analyst should try to establish which characteristics can be modelled independently of the others and which can not. This is important because the modelling of a characteristic can be complex and could require a long process of trial and error. In other to run these processes in a exhaustive way it is very useful to be able to study a characteristic separately of the others. It must be noticed, as it would be explained later, that the estimation process which will be proposed is not one of independent steps but an strategy of estimation in cascade.

To determine quasi-orthogonalities is important to use information from experts. If this information is not available or in order to confirm it, preliminary estimations can be used. In any case, it has been proved useful to start the process considering that the characteristics are orthogonal and pass to model them following the priority ordering stablished according to principle (b).

Modelling in cascade means that one starts by modelling the characteristic ordered in the first place, then eliminating its effects from the data one goes to the modelling of the second characteristic. In a subsequent step both characteristics are estimated jointly, their effects eliminated from the data and one can pass to modelling the third characteristic. Once this is done the analyst could proceed to the joint modelling and the subsequent elimination from the data of the effects of all those characteristics. The procedure would continue till the modelling of the last characteristic and the joint estimation of all of them with a subsequent testing process.

Very often the trend is the most important characteristic of the data and can be considered independently from seasonalities. All this makes that in many cases the characteristics of daily series can be ordered according to the four levels described at the beginning of this section. In that case it only remains to order the different seasonalities. Nevertheless the analyst working with a specific data set should check that such order is appropriate for the data under study and in the negative case should obtain an alternative one

applying the variance reduction criterium.

The other two principles -(c) and (d)- indicate that if there exists indicators (explanatory variables) for specific characteristics of the data one should use them and if not the analyst should propose, by studying the nature of the data, a stochastic or deterministic scheme to model each one of them.

III. Some examples of daily series

Figures 1 and 2 show two Spanish daily series: (1) electricity consumption in Spain excluding the islands and (2) notes and coin in circulation. The simple inspection of these figures reveals that these series have trends and several seasonalities, one of which is annual.

Figures 3 and 4 reproduce the same series for a shorter time span in order to appreciate the interannual seasonalities. For data on stock variables the charts for the levels could conceal some seasonality. This is what happen for the weekly seasonality in figure 4.

Figures 5 and 6 shows the weekly cycle in the electricity consumption and how it is distorted by a midweek holiday. Figure 7 shows the monthly cycles for the data on notes and coin and the alteration suffered by a period of bank strikes in February 1979.

All these figures give an idea of the regularities present in daily series of economic activity and the distortions that these regularities can suffer. The figures also shows how big the seasonal oscillations around the trend can be. The oscillations of the differenced data can be enormous as it is clear from figures 8 and 9.

From all the above information the reader can get an idea of the potential and complexity of modelling daily series. In particular these data show the strong regular patterns in trend and seasonalities that daily series of economic activity have. This means that a model exploiting these regularities would be very rewarding. These charts also point out that the mentioned patterns suffer important alterations, but they are due to specific factors and

therefore are predictable. This implies that a model could capture the regular patterns and the alterations and in that case the model would become an essential tool for forecasting and control the corresponding daily activity.

IV. The use of the variance reduction criterium in daily modelling

The trend and the different seasonalities can be modelled employing: (a) deterministic schemes (dummy variables) or (b) stochastic ones as function of past data values. How one can implement these schemes will be considered in the next section, and the remaining of this section will be devoted to the question of how to choose an alternative versus the other. For that purpose it would be denoted by $T(D)$ and $T(E)$ the deterministic and stochastic schemes for trend and $SJ(D)$ and $SJ(E)$ the deterministic and stochastic schemes for the different J seasonalities, for instance, $J=1$, weekly, $J=2$, monthly and $J=3$, annual.

FIGURE 1.— SPANISH DAILY DEMAND

(Demand for electricity in 1983–1989)

MWh x 1000

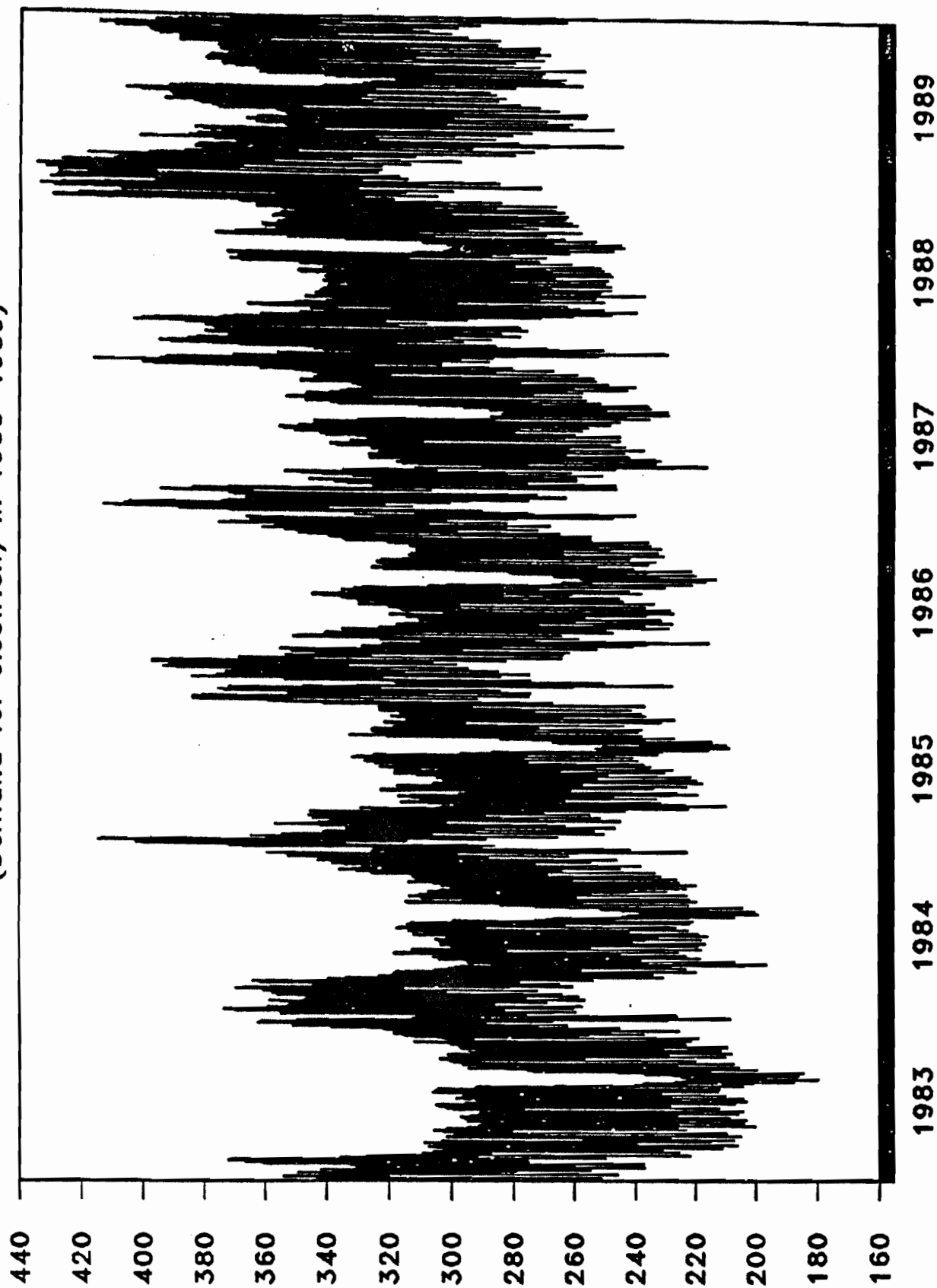
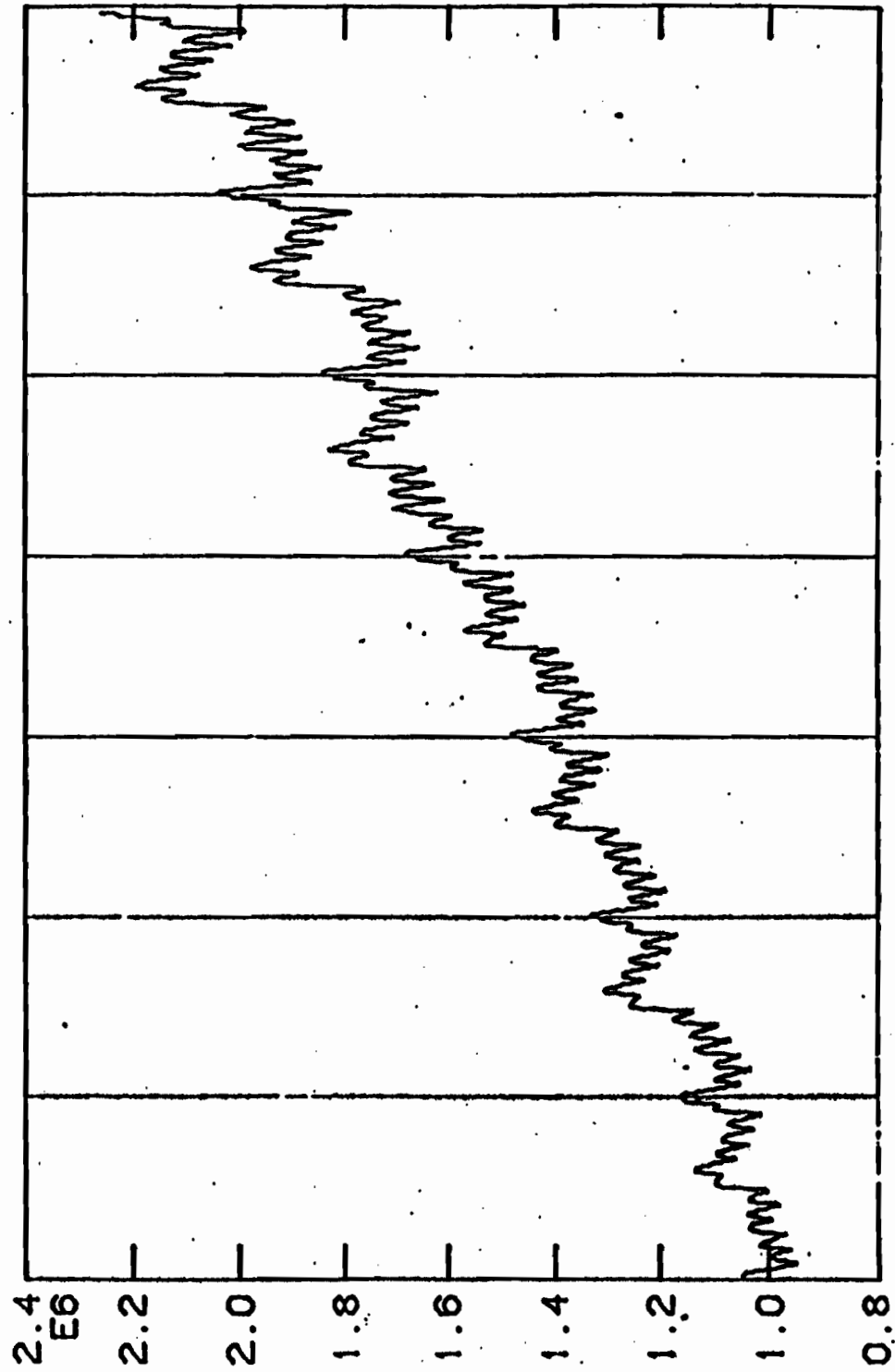


FIGURE 2

NOTES AND COIN (PESETAS) IN CIRCULATION

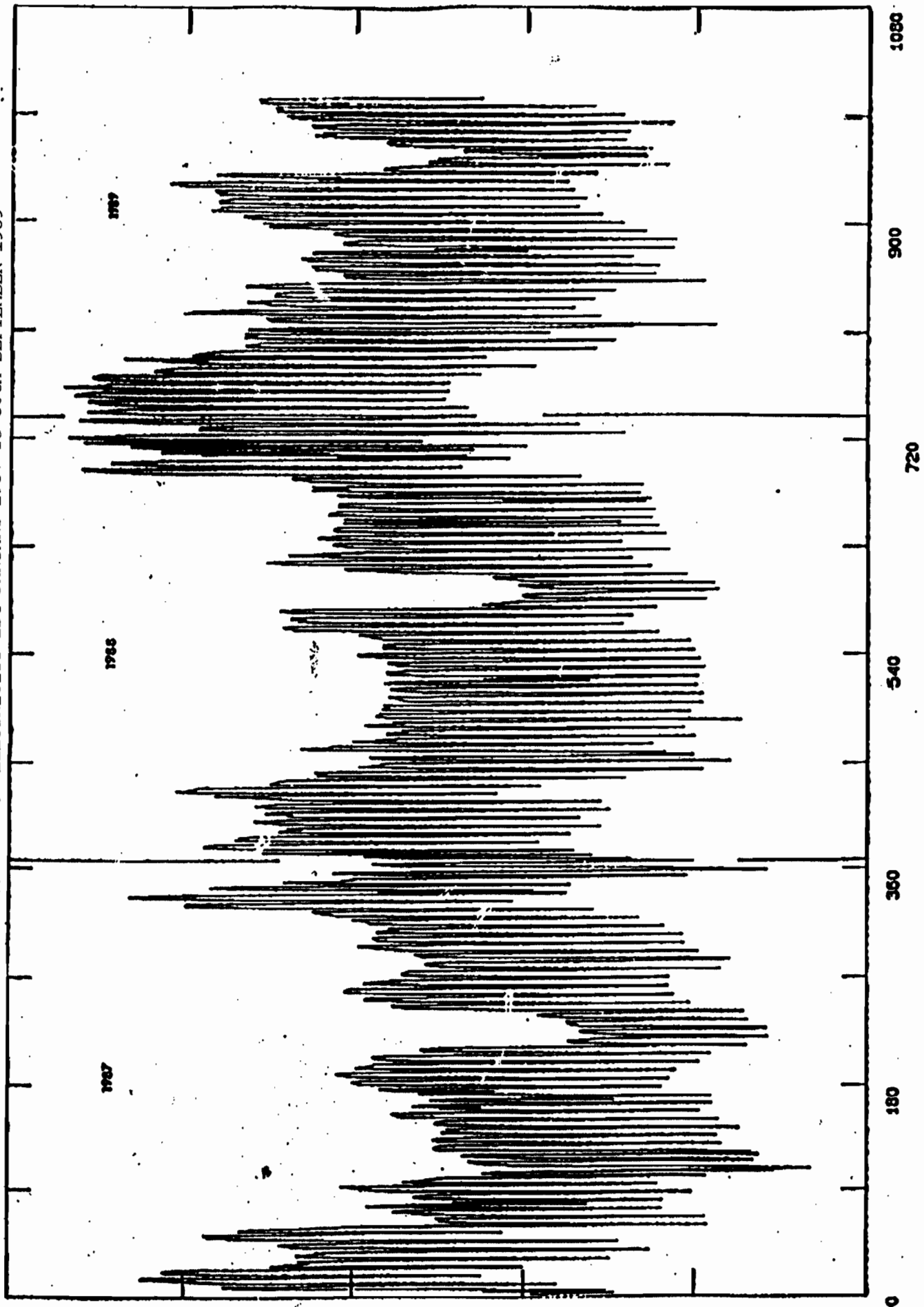


JANUARY 1979

DECEMBER 1985

FIGURE 3

SPANISH DAILY DEMAND FOR ELECTRICITY 1st JANUARY 1987 TO 30th SEPTEMBER 1989



NOTES AND COIN (PESETAS) IN CIRCULATION

FIGURE 4

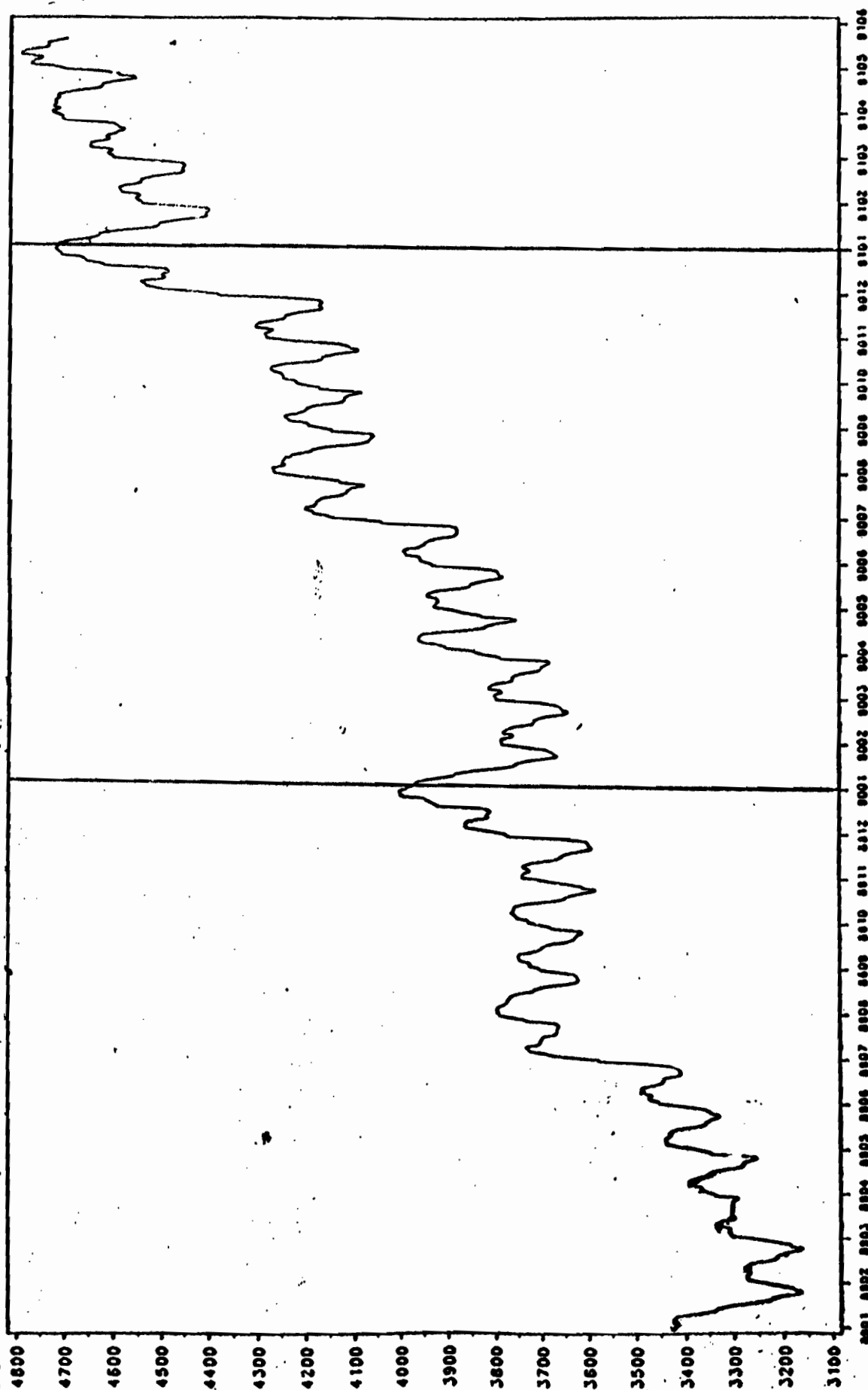


FIGURE 5.- FOUR-WEEK PERIOD DEMAND
(Monday Sep 4 to Sunday Oct 1, 1989)

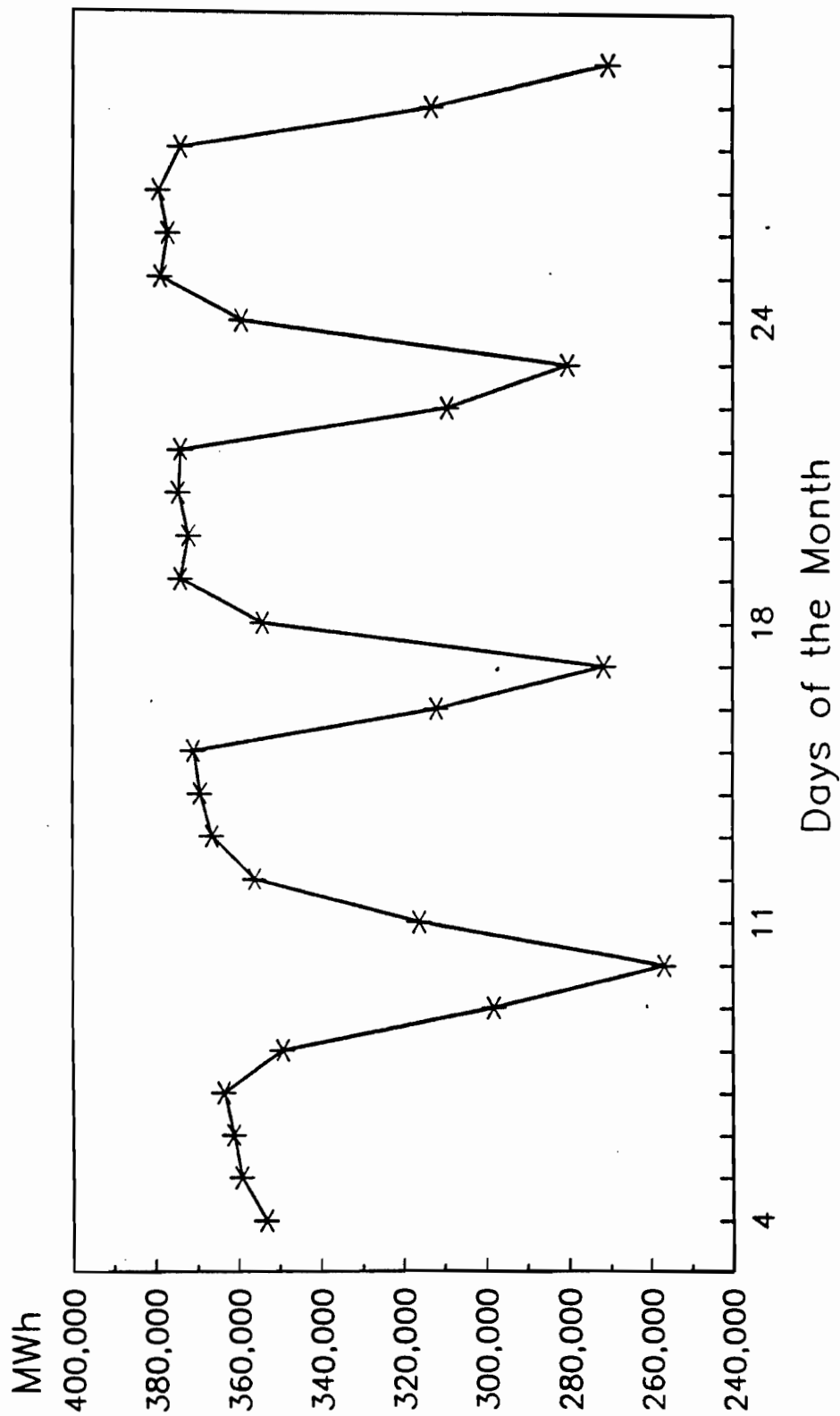


FIGURE 6.- DISTORTION DUE TO A HOLIDAY
(Monday Oct 2 to Sunday Oct 29, 1989)

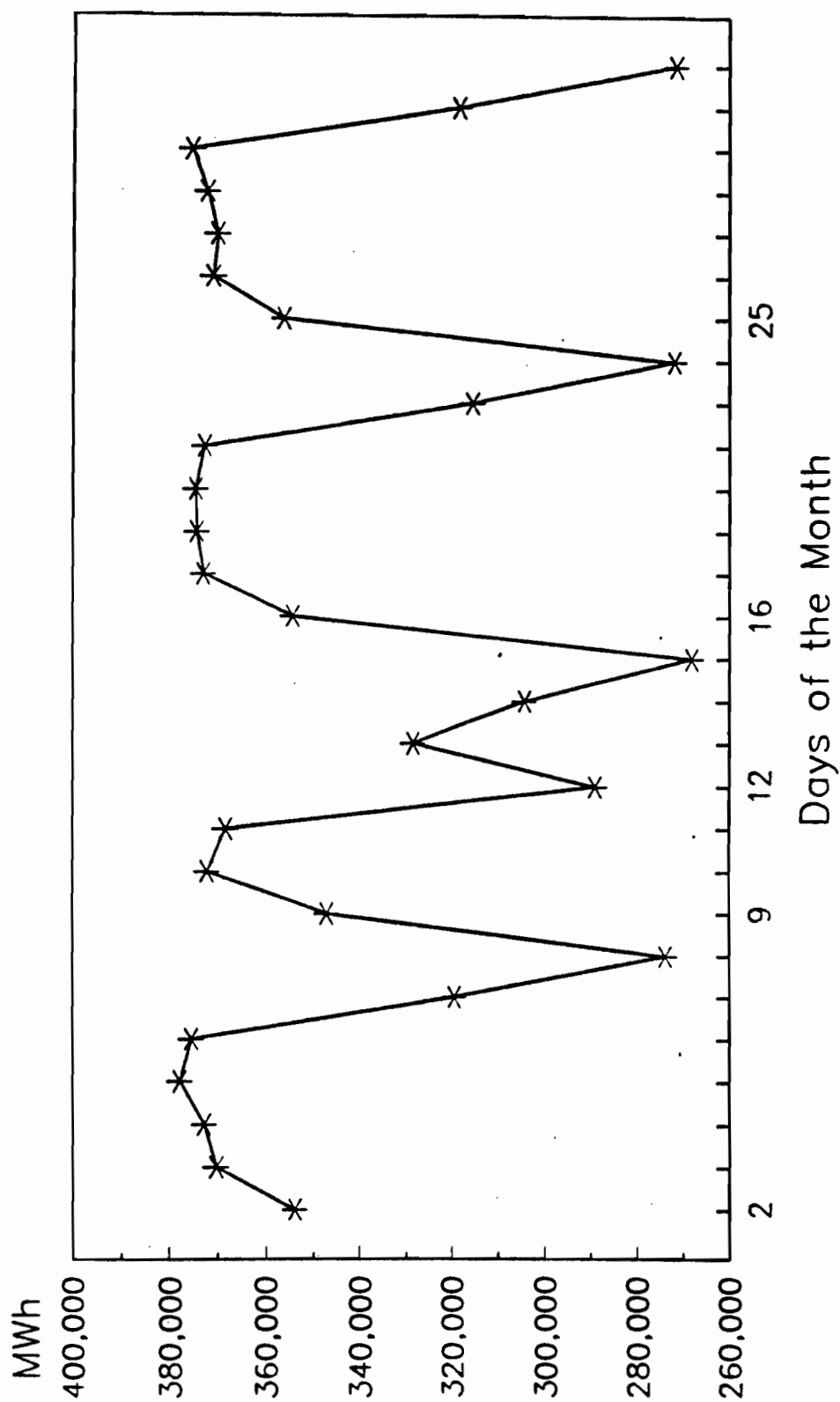


FIGURE 7

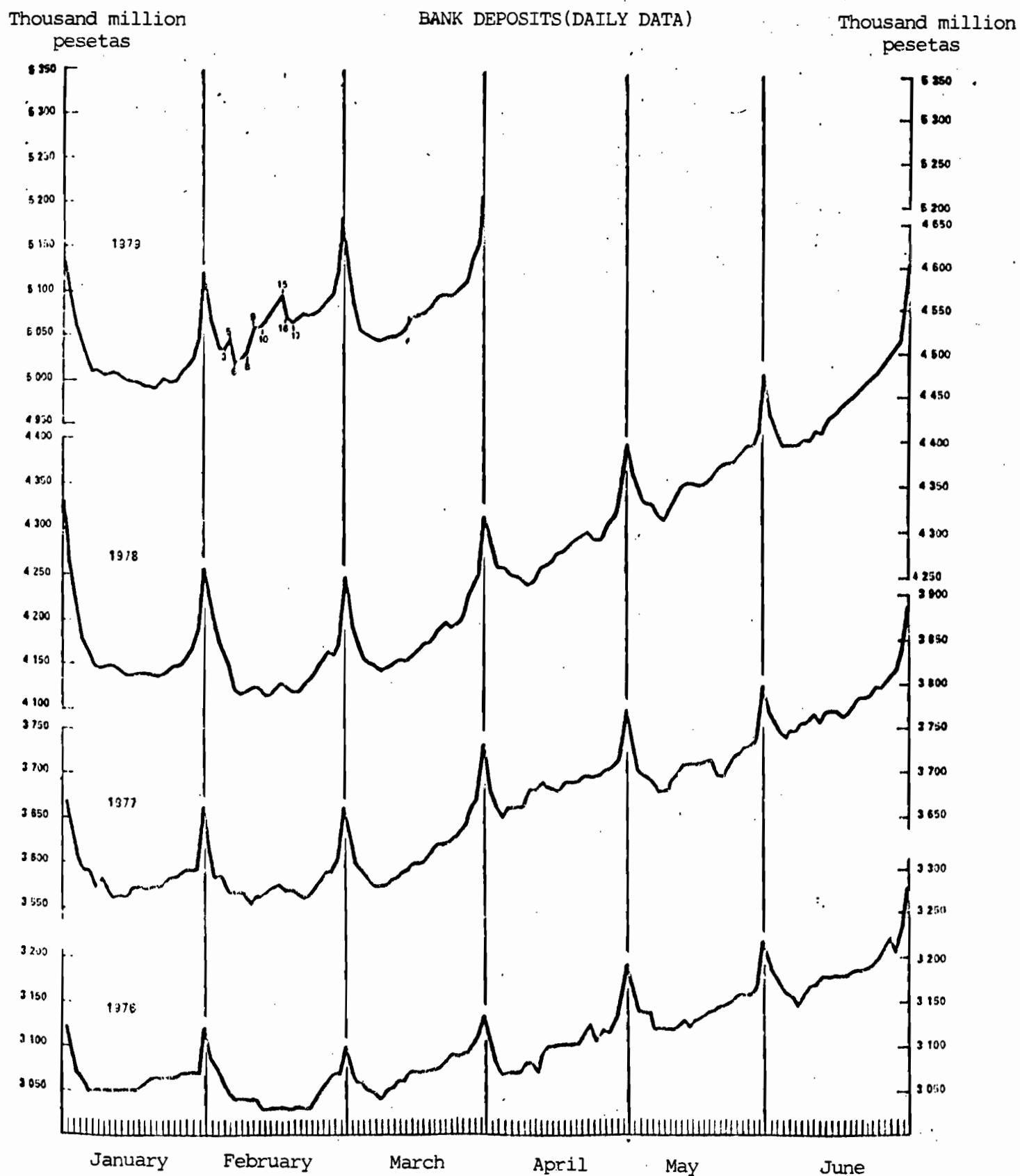


FIGURE 7

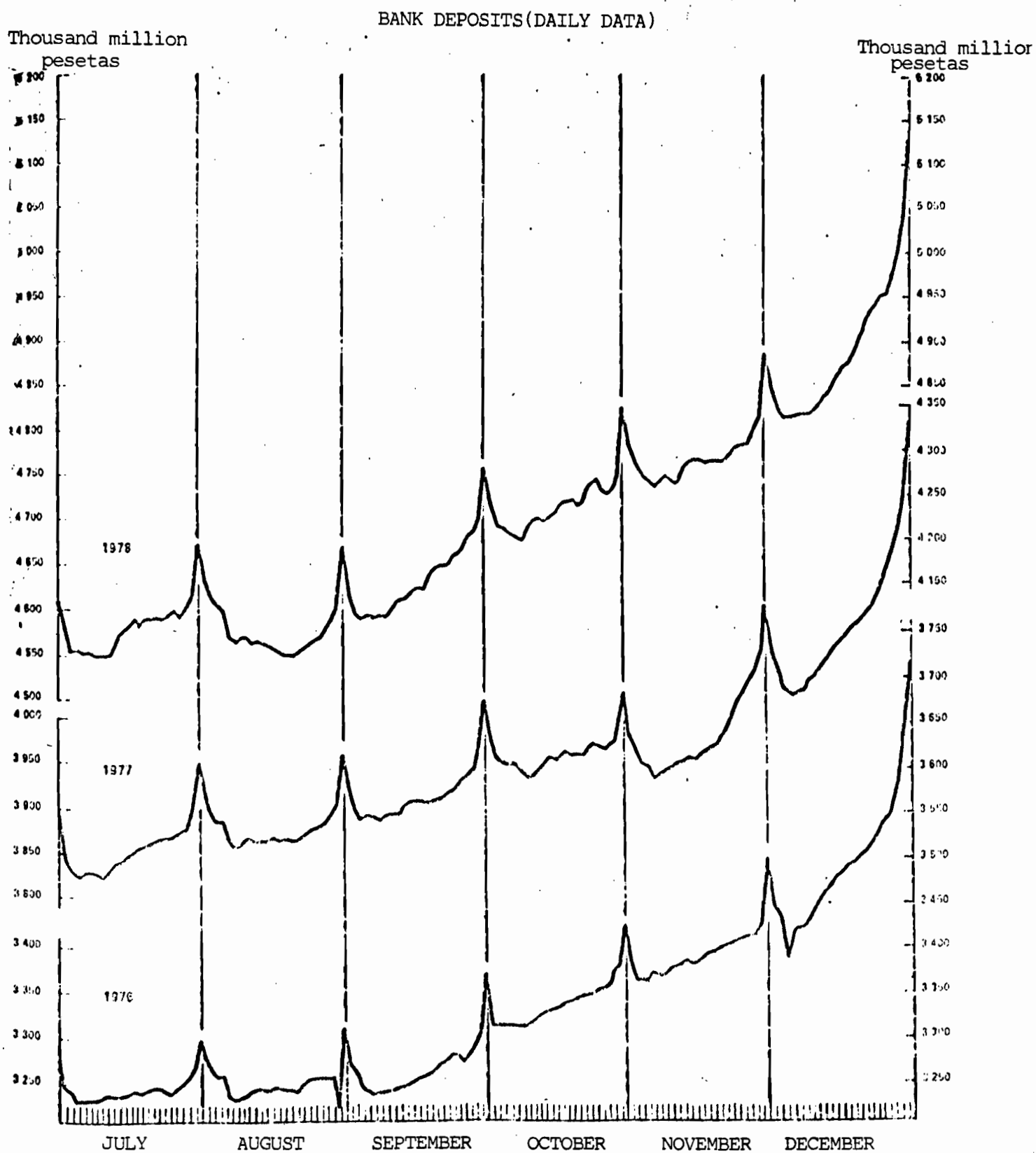
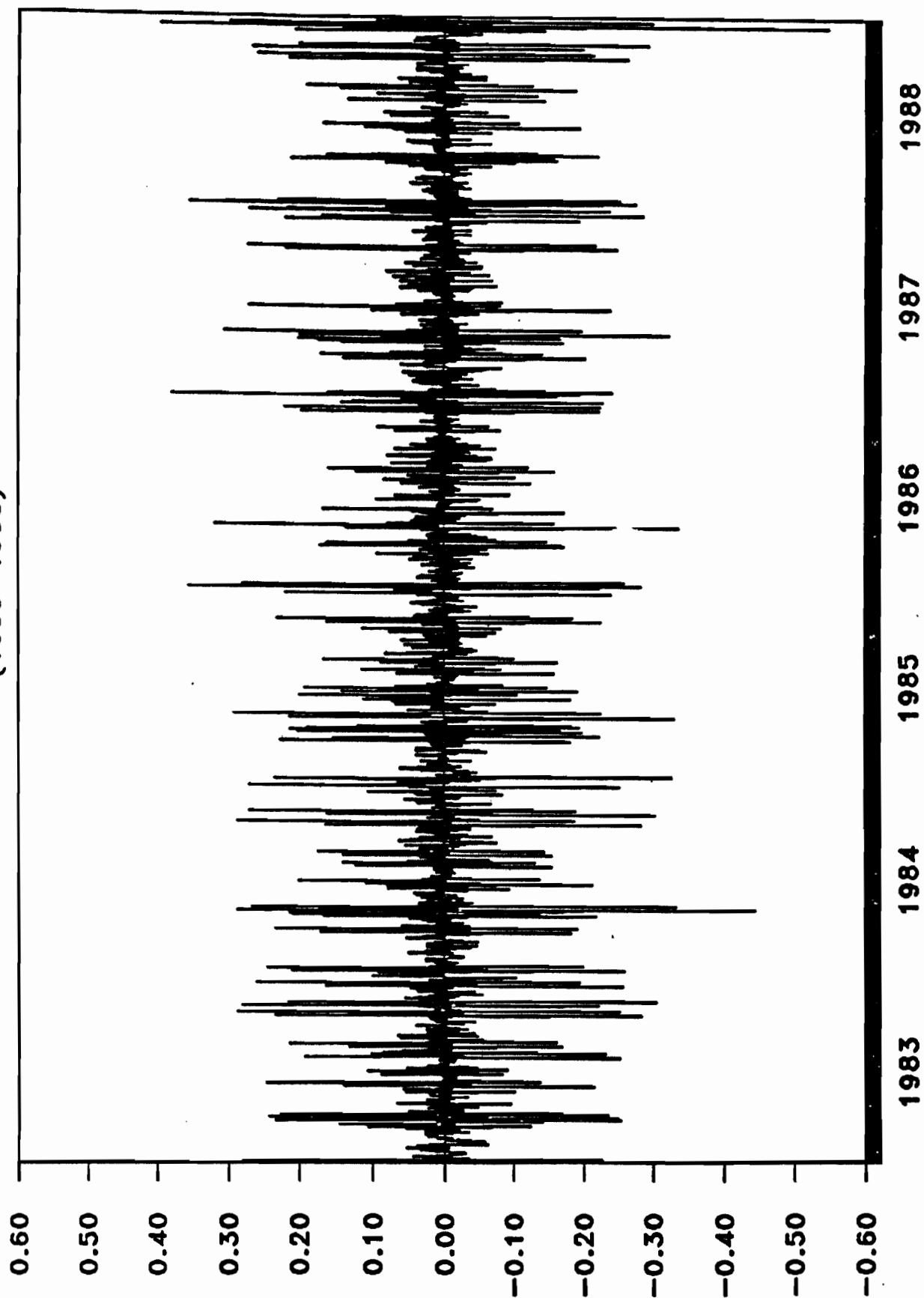
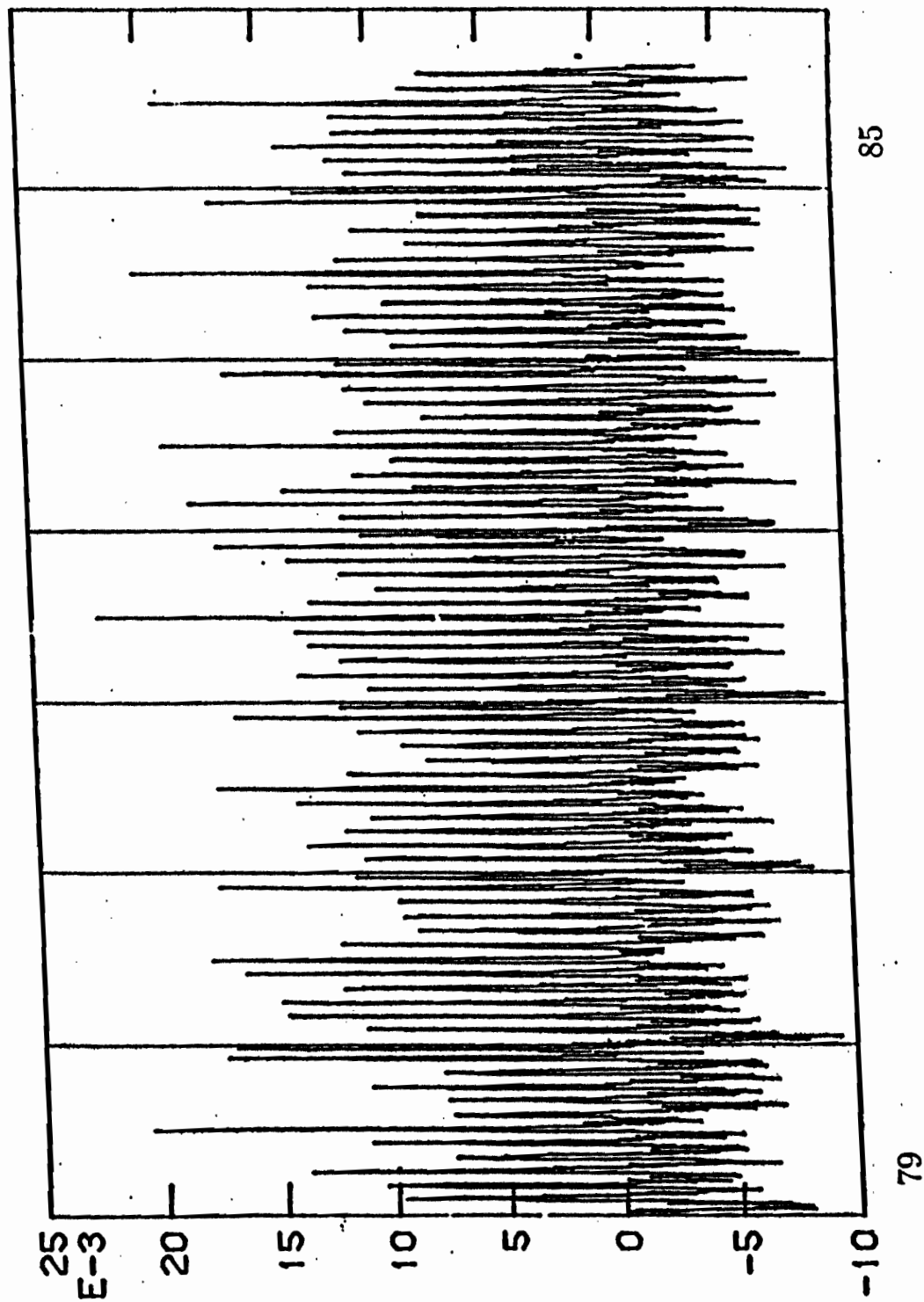


FIGURE 8.— DIFFERENCED SERIES ($\Delta\Delta_7 \ln X$)
 X = Daily electricity consumption in Spain
(1983–1988)



FIRST DIFFERENCES OF X
(X = LOGARITHMIC TRANSFORMATION OF THE NOTES AND COIN (PESETAS) IN CIRCULATION)



In economic time series it is unusual to obtain good fits of deterministic trends when one works with data at a low time aggregation level like a month, week, day.... Therefore for daily series of even a few number of years long it is an exception to find that the deterministic trends are operative and in practice they can be ignored for these data. But for the sake of generality deterministic trends will be included in the discussion.

On the other hand, with regard to the seasonal oscillations of the daily series very often one finds that some seasonalities are better explained by deterministic schemes than by stochastic ones. This could be due to the fact that the stochastic procedures considered apply summation operators over the period of the corresponding seasonality. This implies that for a particular seasonality the analyst is employing a filter with the specific periodicity of the seasonal in question and also a set of filters with all the corresponding harmonic periodicities.

For monthly series the summation operator $(1 + L + L^2 + \dots + L^{11})$, where L is the lag operator, is used to eliminate the annual cyclical oscillations, but also affects the oscillations with period of 6, 4, 3, 2, and 2 months. For monthly series the sum operator usually gives better results than the filter, $1 - \sqrt{3} L - L^2$, specific to remove annual oscillations only. But in daily series, perhaps because we have much more data points, the situation could be different. Thus in comparing the stochastic seasonal scheme -sum operator- with respect to a deterministic one, it could happen that the former perturbs in an unduly way the harmonic frequencies, providing a worse result than a set of dummies. The appropriate seasonal filters for daily data is a topic of much interest which deserves further study. In this paper the only seasonal filter considered is the sum operator.

The trend and the seasonal oscillations in the level of time series are non-stationary factors, therefore if they are appropriately removed the sample variance must be reduced. In view of that if we denote RTD to the residuals of a daily series obtained from the application of a deterministic trend and RTE to the appropriate difference series, one can obtain the sample variances of the original series, X , and of RTD and RTE and calculate the reduction in variance that RTD and RTE amounts with respect to X and one can choose as the way to model the trend the scheme which produces more reduction.

Once the trend scheme has been selected, in what follows it would be denoted by TA, one could pass to the selection of the seasonal schemes. In order to illustrate the process suppose that data has three seasonal factors, S1, S2 and S3, and denote RTASJD and RTASJE the residuals obtained by applying to X the selected trend scheme (TA) and the deterministic (D) or stochastic (E) schemes corresponding to one of the SJ (J=1, 2, 3) seasonalities. By obtaining the sample variances of RTAS1D, RTAS1E, RTAS2D, RTAS2E, RTAS3D and RTAS3E it can be seen which seasonal scheme implies more variance reduction with respect to RTA and this scheme will be chosen as the appropriate one to model to corresponding seasonality. Suppose that this is S1 and denote by A the selected scheme. Then one can obtain the residuals from models for X with schemes TA, S1A and one of the followings: S2D, S2E, S3D and S3E. The corresponding sample variances can be calculated and it would be chosen as the adequate new scheme the one which implies more reduction with respect to the variance of RTAS1A. Denoting by S2A this selected scheme and by RTAS1A S2A the corresponding residuals one can continue to obtain the residuals from models including, TA, S1A and S2A and one of the following schemes: S3D and S3E. Again it would be chosen the scheme with more variance reduction with respect to RTAS1AS2A.

If in a particular step of the application of the previous procedure the inclusion of TA, S1A, ... does not suppose a variance reduction with respect to the residuals obtained in the previous step, it would be taken as indication that the seasonalities not included in the previous step are not present in the data or they are such that require to be modeled by using some indicator variables. This ambiguity will be solved using prior knowledge about the data. If there are reasons to suppose that a seasonality is present but its modelling requires the use of indicators, usually these can be found using meteorological variables or a specific type of dummies based on prior knowledge.

The application of the principle of variance reduction for cases in which two characteristics are far from being orthogonal is straight forward. It only requires to estimate them jointly when comparing sample variances.

IV. Modelling the trend and the seasonals without indicators.

V.1. Modelling local oscillations and trends

In general, when modelling daily trends the analyst finds that there is no indicator for them. Thus the daily interpolation of the industrial production index or of any other monthly indicator that could be related with the daily data under study turn to be not very useful.

The representation of the model by a deterministic function of time, for instance a linear trend like

$$X_t = a + bt + R_t,$$

is excessively rigid and the corresponding residuals, R_t in the example, usually have a nonstationary level.

The previous considerations suggest that daily trends could be modelled stochastically as a function of previous data. Thus for series with persistence in their levels, for instance pollution and traffic levels, some transport data, etc., the model could be:

$$X_t = \underbrace{X_{t-1}}_{\substack{\text{local level} \\ \text{(persistence)}}} + R_t,$$

which implies to apply first differences to take account of the stochastic persistence in the level of the data.

For series with accumulation in their levels the model for the stochastic trend will be:

$$X_t = \underbrace{X_{t-1}}_{\text{local level}} + \underbrace{(X_{t-1} - X_{t-2})}_{\substack{\text{local increment} \\ \text{(accumulation)}}} + R_t,$$

which implies the application of double differencing.

V.2. Modelling seasonalities by deterministic schemes.

For the deterministic modelling of a certain seasonal factor the analyst can use dummy variables. Thus, in order to model the seasonal factor SJ with periodicity P_j , the dummies employed would be $VAJ(h)$, $h = 1, 2, \dots, P_j$, where $VAJ(h)$ has value one in those observations corresponding to the seasonal moment h of the seasonal factor SJ and zeros everywhere else.

The monthly seasonality will be dealt specifically later on. But at this stage one could say that for this seasonality P_j will be taken as 30, with dummies $VA30(1)$, $VA30(2)$, ..., $VA30(30)$. For months with 31 days the variable $VA30(15)$ will take value one for the days 15 and 16 of these months and the remaining dummies will be assigned correspondingly to the days of the month in question. For the months of February with 28 days $VA30(15)$ and $VA30(16)$ will take values zero and $VA30(17)$, $VA30(18)$, ..., $VA30(30)$ would take value one for the days 15th, 16th, ..., 30th, respectively, and for leap years only variable $VA30(15)$ will be omitted.

For the annual seasonality it is not at all advisable to employ freely 365 daily dummies. The deterministic treatment of this seasonality must be applied restricting the coefficients of these daily dummies to be the same for days belonging to the same week, fortnight or same month.

For each set of seasonal dummies employed their corresponding coefficients must be restricted to sum zero, in order to peak up a pure seasonal effect. Of course, the failure to introduce this restriction will induce multicollinearity with the trend.

In some daily series the analyst would find that seasonal factors of short period, like a week, can change along the year, for instance, in the summer holidays. In this case two sets of weekly seasonal dummies would be required, one for each regime. This has a natural extension to cases with more than two regimes along the year.

V.3. Stochastic modelling of the seasonals.

The stochastic modelling of a seasonality of periodicity P_j will be done by applying the sum operator: $U_{p_j}(L) = (1 + L + L^2 + \dots + L^{P_j-1})$ to the original series.

If besides this operator the series requires differencing to take account of the trend then the combination of $U_{p_j}(L)$ with a difference operator implies a seasonal differencing factor, i.e.

$$(1-L) U_{p_j}(L) = (1 - L^{P_j}).$$

In this case the model for stochastic trend and seasonality would be

$$X_t = \underbrace{X_{t-p_j}}_{\text{seasonal level}} + \underbrace{(X_{t-1} - X_{t-1-p_j})}_{\text{annual increment}} + R_t$$

VI. Meteorological variables as indicators for annual seasonality

The annual seasonality can be mainly due : (a) to social habits, as it uses to happen with data for monetary aggregates, sales, some transport series, etc, or (b) to meteorological variables, like is the case of consumption in electricity, water, etc. In the first case one can use the methods, well deterministic well stochastic, described in the previous section, because it is pretty difficult to reflect the evolution of the factors causing this type of seasonality in form of real quantitative indicators. In the second case these indicators are available: meteorological variables, and this section is devoted to discuss how to face the problem of annual seasonality when it strongly depends on meteorological variables.

In such cases there are a set of questions that must be taken into consideration:

- (a) The meteorological variables usually would not explain all the annual seasonality in the data.
- (b) Temperature is generally responsible for the annual oscillations in the daily data, but

it is not necessarily the only meteorological variable contributing to this seasonality.

- (c) To capture the seasonal effect the analyst, very often, must transform the meteorological variables.
- (d) The effect of meteorological variables can change with the type of day.
- (e) The meteorological variables not only contribute to explain the annual seasonality, but also their short-term oscillations affect, sometimes in a very notorious way, the short-term of the daily series.
- (f) The effects of the meteorological variables can be nonlinear.
- (g) The influence of the meteorological variables can be captured more precisely using dynamic response functions.

As it has been mentioned in point (e) the consideration of meteorological variables is useful to capture some seasonal component in the data and to explain irregular or short-term oscillations, but it does not seem advisable to decompose these explanatory variables in two components (signals) to explain these effects.

This is so because such seasonal and irregular features in daily data can have a nonlinear and dynamic structure, and as a consequence of that the extraction of seasonal and short-term signals from meteorological variables to use them separately in a daily model could result in a very difficult modelling task. In fact, to extract signals one would use dynamic filters, but if the signals obtained do not approximate relatively well the seasonal and non-seasonal signals to which the agents react, the dynamic response functions that the model could require for those signals could be difficult to obtain. Besides that the agents could have similar reaction functions to both signals, which it would make irrelevant the disaggregation for the purposes of the daily model. For all these reasons it is not a bad recommendation, at least for the time being, to use each meteorological variable as a whole in the modelling process. Once the model has been constructed the observed values of the meteorological variables, for instance temperature, can be split in two additive parts, the "normal value" for corresponding day of the year and a "deviation" for it, and use the model to evaluate the effect of temperatures which are deviations from normal values.

The possible nonlinear and dynamic effects of the meteorological variables are in a

great deal related also with the short-term oscillations in the data and it will be considered later. In this section the discussion will be centered around questions which are more specific for the annual seasonality.

When the meteorological variables have big power to explain the annual seasonality in a daily series the analyst usually finds that social habits which can not be approximated by these variables also have some influence in the annual seasonality. Thus the working calendar and the vacation periods must be taken into consideration. Figure 10 which represents the electricity consumption in Spain in a specific year, shows how this consumption falls in Easter, August and Christmas. That these falls can not be explained by meteorological variables it will be clearer later on.

Temperature uses to be the most relevant variable to explain the annual seasonality in daily series and it has a strong annual cycle. Nevertheless temperature oscillates a lot around it and an estimation of its cycle could be useful for a better understanding of the relation between temperature and the daily variable under consideration. In particular it can be useful to learn the type of transformation which temperature could required before entering into the daily model.

The annual cycle for a temperature series can be obtained by applying a centered moving average to the series, splitting the resulting values by years and averaging the values of each day along years. Applying this procedure to a weighted average of daily temperatures in different parts of Spain one gets an aggregate "normal temperature" for Spain along the year. This result is shown in figure 11.

Once the analyst has an estimation of the annual cycle in the relevant temperature series it is convenient to compare it with a rough estimation of the annual cycle of the daily data. This can be obtained, for instance, calculating the difference between each monthly average and the total average. The thick line in figure 12 represents the estimation of the annual cycle for the electricity consumption in Spain. In this figure it can be seen that August supposes a break over an smooth yearly cycle. A rough correction for the behaviour of this consumption due to summer holidays, in August, according the discontinuous line in figure

FIGURE 10.— DAILY DEMAND FOR 1989

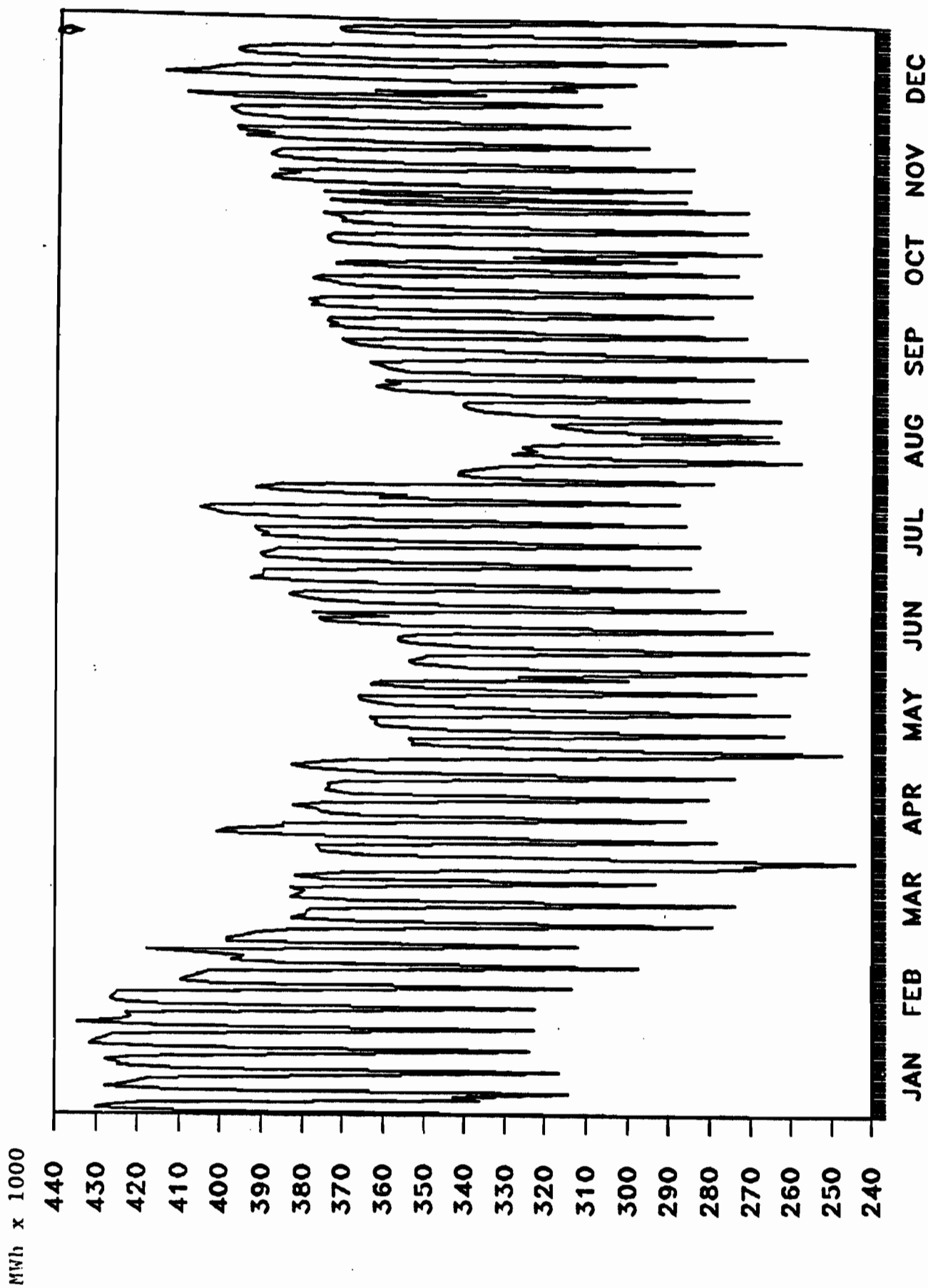


FIGURE 11

ANNUAL CYCLE OF TEMPERATURE IN SPAIN

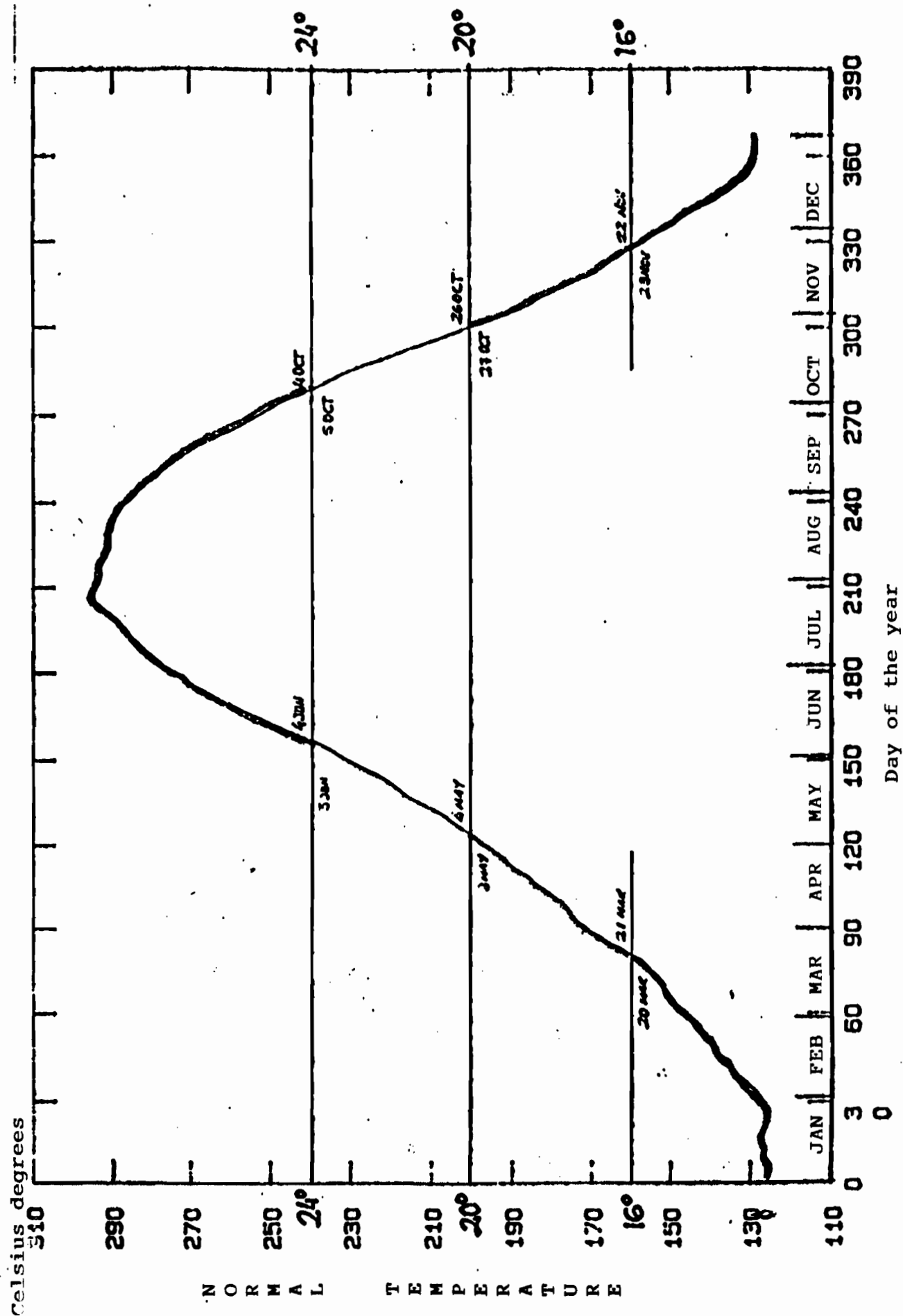
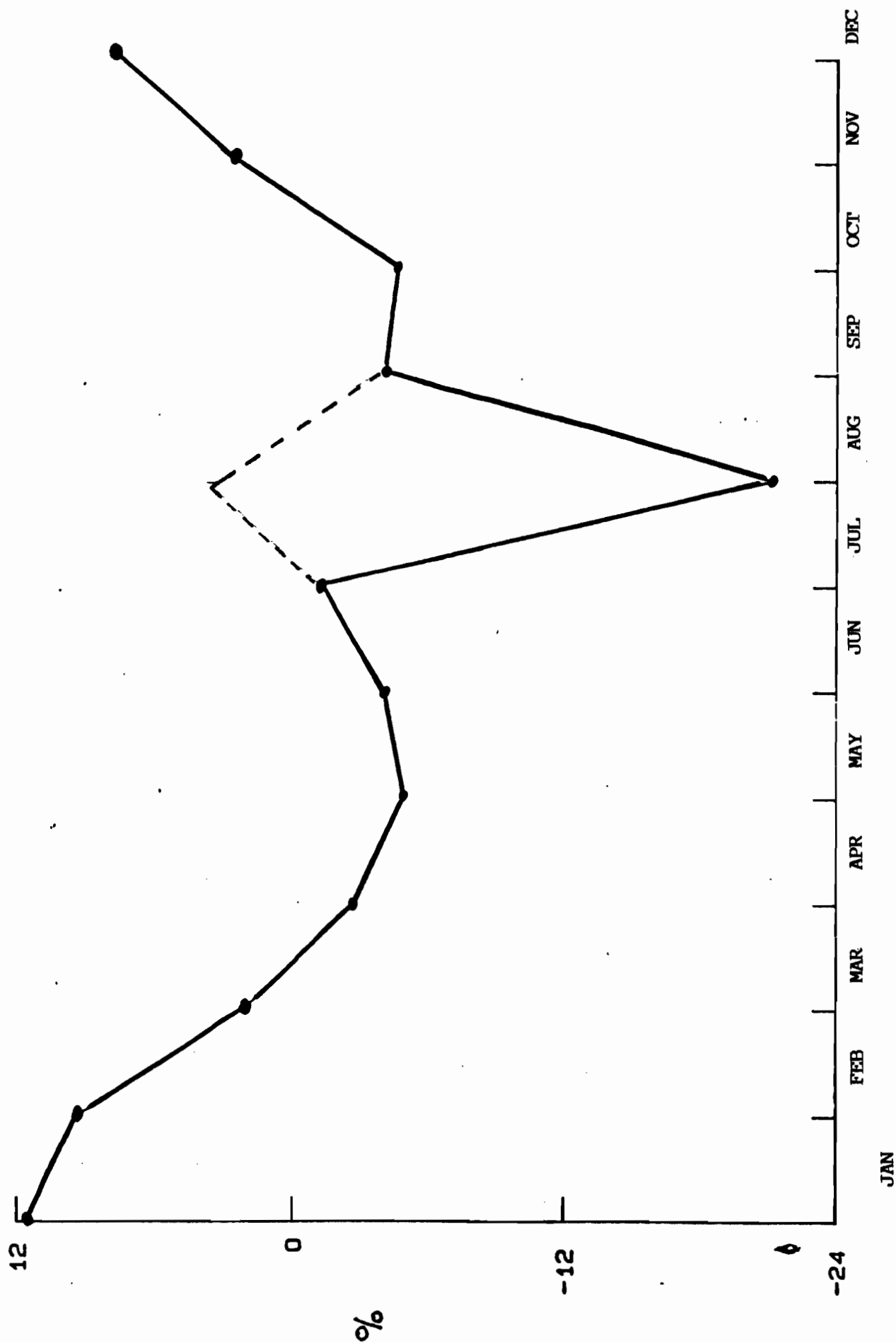


FIGURE 12

Percentage of the monthly means over the total mean in Spanish electricity consumption



12, would lead to a version of some sort of corrected annual cycle.

Following with this example one has that comparing the corrected cycle in consumption with the cycle in the temperature data there are two aspects that results of some interest in order to find and appropriate transformation of temperature for the problem in question:

- (a) for low temperatures the relationships with consumption is negative, but with high temperatures the relationship is positive;
- (b) the elasticity of consumption with respect to temperature is, in absolute values, greater for low temperatures than for high ones.

All this has two interesting implications. One is that there is some sort of "comfortable band of temperatures" such that below it we could say that temperatures are "low" and above it that they there are "high". The other is that the low and high temperature values must be enter into the model as separate explanatory variables in order to estimate different elasticities.

This example shows how the meteorological variables could need transformation before to enter them into a daily model. The very precise way in which the temperature values must be transformed could be something uncertain, in which case the analyst should prove with different reasonable alternatives and choose the one which provides a better fit to the data. For the previous example Cancelo y Espasa (1991) have proved with different values to define the comfortable temperature band and found that 20 y 24 centigrade degrees provide the best fit.

The effects of meteorology in daily data could be different depending the type of the day. The types of days that one can have in a particular problem can vary from one case to another but some classifications that appear useful in cases are:

- (i) working days and holidays and weekends.
- (ii) end and beginning-of-month days and the rest.

- (iii) days with summer working hours and the rest
- (iv) one type of day for each season in the year, etc.

Some daily data require to specify types of days crossing two or more of the above mentioned criteria, for instance, crossing (i) and (iv)

The different type of days can be captured by binary dummies and multiplying these dummies by the corresponding meteorological variable, say temperature, one obtain the different temperature variables that should be used as explanatory ones in the daily model.

VII. End and beginning-of-month effect.

In monetary aggregates, see figures 2, 4 and 7, but certainly in many other daily series like some transport data, sales of certain goods, etc. one finds that data shows oscillations which its periodicity is one month. This is usually due to the fact that salaries are generally paid at the end of the month.

In principle one could attempt to capture this cyclical behaviour by a sum operator. Since a month has a different number of days along the year the analyst should face the problem of fixing the order of the sum operator. One could consider that the average number of days in a month is 30.44 and the median is 31, therefore could choose a sum operator of 30 or 31 days. But both produces distortions in the data. These distortions in cases are important and then one should considered and alternative way to proceed.

On the other hand the end and beginnig-of-month effect of a particular month tends to be more similar to the same month of the previous year than to previous months. The result is that the end and beginning-of-month effect generates some sort of annual cycle which could be considered in the model by a sum operator of 365 days. Since an harmonic of the annual cycle is the cycle of 30.33 days one has that the annual sum operator has anyway some power to explain a monthly cycle. To proceed in this discussion we will assume that besides the annual sum operator the data require at least one difference and

therefore the combination of both results in an annual difference operator $(1-L^{365})$. If there were no leap years the annual difference would transform the daily series of seven days per week on daily increments over the same day in the previous year. But the presence of leap years and the fact that in many problems, certainly money aggregates, the daily series are of only six or five days per week, the corresponding annual difference does not always produces annual increments over exactly the same day of the previous year. This is an important distortion for the end and beginning-of-month effects. These distortions can be captured by dummies.

Table one shows the months between August 1988 and December 1990 for which the annual differences in daily series with five days per week produces distortions in capturing the effect and therefore the months which require the presence of a dummy.

Very often series showing a strong end and beginning-of-month effect have different weekly cycles depending whether the week is in the transition from a month to another or not. In those cases the analyst can model the weekly cycles by using two types of weekly dummies, one for weeks which belong only to one month and another for weeks belonging to two months.

The end and beginning-of-month dummy variables would affect several days and they should be included in the model with a dynamic filter which could be different for different months or for months corresponding to different seasons of the year. Also this filter could be different when Easter is at the end of March or beginning of April. The precise design of all these effects is something that the analyst should discover by trial and error using different alternatives. At the end these effects are estimated with very few degrees of freedom but the estimates turn to be quite reliable. The problem is that with samples of just a few number of years some distortions at the end and beginning of a month could not have happened previously (see table one) and for forecasting propurses should be treated as one of the estimated ones which could be similar to it. This implies the necessity of estimating the daily model each time that a new distortion at the end and beginning of the month appear.

TABLE 1

End and beginning-of-month distortions with an annual difference operator in daily series of five observations per week.

August 1998 - December 1990

	1988	1989	1990
JAN-FEB	-	YES	NO
FEB-MAR	-	NO	NO
MAR-APR	-	NO	YES
APR-MAY	-	YES	NO
MAY-JUN	-	NO	NO
JUN-JUL	-	NO	YES
JUL-AUG	-	NO	NO
AUG-SEP	YES	NO	NO
SEP-OCT	YES	YES	YES
OCT-NOV	NO	NO	NO
NOV-DEC	YES	NO	NO
DEC-JAN	NO	YES	NO

VIII. Some examples in modelling trend and seasonalities in daily series

As an illustration of the procedure described previously the two series presented in section III will be used here. The results of the sample variance reduction criterium applied to the electricity consumption (D) are collected in table 2. For this daily series deterministic schemes to capture the different seasonalities did not produce good results and in order to simplify the table they have been omitted from it. In the construction of this table difference operators have been used to take account of the trend in the data and the sum operators - defined in section V- $U_7(L)$, $U_{30}(L)$ and $U_{365}(L)$ to take account of the possible weekly, monthly and annual seasonalities.

Table 2 points out three main results:

- (a) The trend and the weekly seasonality are the more relevant features of these data since the greatest reduction is obtained with $(1-L)(1-L^7)$ operator.
- (b) The monthly seasonality does not seem to be present in the data since the sample variance of $\Delta_7 \Delta_{30} \ln D_t$ is of the same order of magnitude that the sample variance of $\Delta \ln D_t$, when the data clearly requires double differencing. This result agrees with the prior information about these data and therefore one does not need to worry any more about monthly seasonality in this series.
- (c) The annual sum operator does not seem convenient because the operator $(1-L^7)(1-L^{365})$ increases the variance with respect to $(1-L)(1-L^7)$. Nevertheless, now the results are different from those in (b) because $(1-L^7)(1-L^{365})$ reduces variance with respect to $(1-L)$, indicating that it could be interesting a further analysis on the annual seasonality. In fact, the prior knowledge on this data indicates that this annual seasonality is there but is mainly due to the important effect that meteorological variables have on electricity consumption and such variables oscillate very much around their annual cycles. The conclusion is that without these indicators the electricity consumption can not be modelled properly.

Table 2

**Sample variance reduction in the daily electricity
consumption in Spain (excluding islands)**

Transformation (*)	Mean (t) (**)	Stand. Dev.	Q=40	Q=380	Q=750 (***)
LD	12.6(3424)	14.78%	7405	40858	68698
Δ LD	.0001(.04)	12.95%	7216	65081	-
Δ_7 LD	.0002(.08)	8.65%	1219	3201	4775
$\Delta\Delta_7$ LD	.0000(-.01)	7.40%	585	2100	3222
$\Delta_7\Delta_{30}$ LD	.000(.01)	12.59%	2690	5825	8393
$\Delta_7\Delta_{365}$ LD	-.0001(-.03)	9.33%	1461	3885	5567

(*) $LX = \ln X$.

(**) Between brackets the t statistics for the mean.

(***) $Q = j$: Ljung-Box statistic calculated with j lags of the correlogram.

Table 3

**Sample variance reduction in the daily series of notes
and coin in circulation in Spain (six observations per week)**

Transformation	Mean	t statistic for the mean	Standard Deviation	Ljung-Box statistic	
				Q=36	Q=350
log CF	14.2005	2899.7	0.1954	73,918	-
Δ log CF	0.0003	2.7163	0.0053	6,034	78,720
Δ^2 log CF	0.0000	0.0192	0.0032	611	6,233

Transformation	Mean	t statistic for the mean	Standard Desviation	Ljung-Box statistic	
				Q=36	Q=350
$\Delta\Delta_6$ log CF	0.0000	0.0568	0.0075	6,164	71316
$\Delta\Delta_{26}$ log CF	0.0001	0.5745	0.0048	1,295	12309
$\Delta\Delta_{313}$ log CF	0.0000	0.0065	0.0028	235	4607
$\Delta_6\Delta_{26}$ log CF	0.0005	1.0349	0.0205	10,105	-
$\Delta_6\Delta_{313}$ log CF	0.0000	0.0399	0.0080	2,339	8373
$\Delta_{26}\Delta_{313}$ log CF	0.0000	0.1143	0.0137	11,326	-

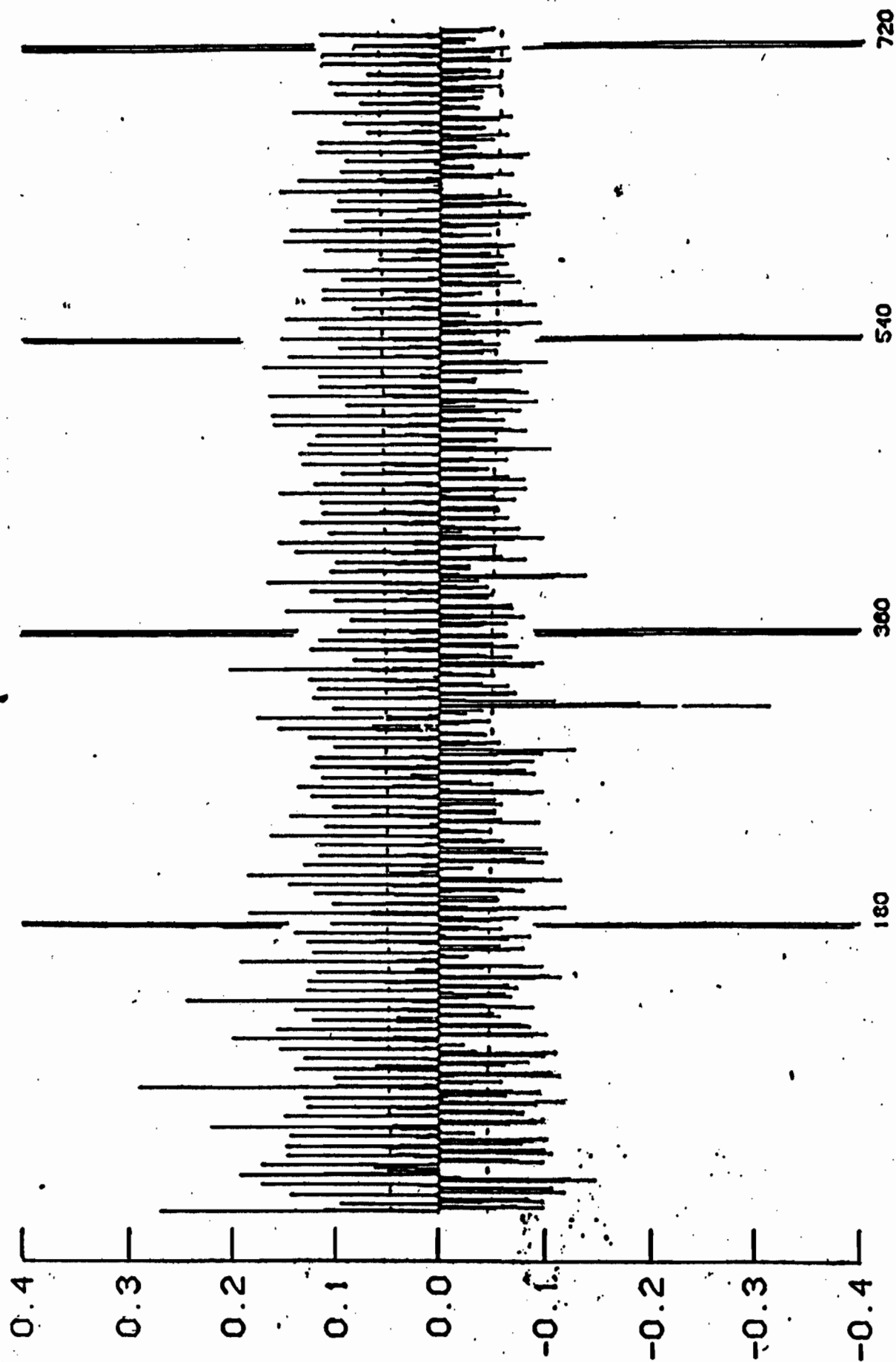
The results of the sample variance reduction criterium on the daily series of notes and coin in circulation in Spain, CF, are in table 3 and they can be summarized as follows:

- (a) The trend and the annual seasonality are the more relevant features of these data since the greatest reduction is obtained with $(1-L)(1-L^{313})$.
- (b) The monthly sum operator does not seem convenient for these data because increase the sample variance with respect $(1-L)(1-L^{313})$. But on the other hand the $(1-L)(1-L^{26})$ reduces variance and simplifies the time dependence -the Ljung-Box statistics have lower values- with respect $(1-L)$. So, for this series, it is very important to analyze if the annual difference operator is distorting the data in the transition of a month to another and in the positive case to take account of it by the use of dummy variables. Prior knowledge and the correlogram of $\Delta\Delta_{313}$ in CF, figure 13, confirm the presence of the end and beginning-of-month effect in this series.
- (c) The fact the variance of $(1-L)(1-L^7)$ in CF_t increases over the variance of $(1-L)$ in CF_t indicates that well the weekly seasonality is not in the data or it requires alternative schemes. Certainly the prior knowledge and the correlogram of $\Delta\Delta_{313}$ in CF_t , figure 13, indicate that the weekly cycle is in this series and also the presence of end and beginning-of-month effects points out the possibility that the weekly cycles could be different at the end of month. In such situation it is worthwhile to study the modelling of the weekly seasonality by the use of two different sets of weekly dummies.

As a summary of these two examples it could be said the following. Both series show a trend and annual and weekly seasonalities, but the first mentioned seasonality in the electricity consumption and the second in the notes and coin in circulation can not be modelled properly as a function of past data. Therefore, only the trend and one seasonal factor -weekly for the electricity consumption and annual for the notes and coin in circulation- can be modelled using autoregressive schemes with unit roots. The annual seasonality in the data for electricity consumption must be modelled employing indicators taken from meteorological variables. In the case of the series of notes and coin in circulation the end and beginning-of-month effect are also present and the weekly cycle changes in the transition from a month to another. All this suggests that these monthly and weekly effects should be modelled by the appropriate design of dummy variables in both cases and by

FIGURE 13

$\Delta\Delta_{313}$ ln CF correlogram



DALCFIN : DALCFIN=DLCFIN-TSLAG(DLCFIN,313)

adding adequate dynamic filters in the case of the end and beginning-of-month effects.

IX. Modelling calendar effects

Alterations in the working calendar usually have effects, sometimes very notorious, in daily series. When these alternatives are present they can be modelled using dummy variables. But the design of these dummies and of the possible dynamic filters that they could require is in cases complex and a systematic approach in modelling calendar effects can simplify very much the task.

In order to construct a useful approach it turns out to be helpful to classify the main factors changing the working calendar as:

- (1) holidays
- (2) vacation periods
- (3) special events

In what follows the procedures presented will be discussed under separate headings for each factor.

IX.1. - Holidays

A holiday other than Sunday in a week can affect the behaviour of a daily series in some complex and important way and the impact of a holiday could differ from the impacts of other holidays for several reasons. Therefore, in the willing of retaining as many degrees of freedom as possible one can not establish that the effect of a holiday at any moment in time is always the same, because this is false in general. A more convincing approach is to separate holidays with specific effects in the year and holidays which can be grouped into categories such that those belonging to the same category have the same effect.

A criterium to group holidays is :(C1) according to the day of the week, but in some cases this criterium only is too restrictive and becomes useful to group holidays employing as a second criterium (C2). This second criteria could be:

- C.2.a. the season or month of the year.
- C.2.b. the first o second fortnight in the month (or the order of the week in the month).
- C.2.c. as a function of meteorological variables.

In C1 one does not need to consider necessarily six different groups: holidays on Mondays, ..., on Saturdays, sometimes two or three groups could be enough. In this C.2.a the definition of season can be adapted to the characteristics of the daily series in question and the number of seasons in the year could be different from four. In C.2.b a month can be divided in a number of parts different than two or four if the problem in question demands it. Finally, in C.2.c the criterium could be holidays with sunshinne or not, holidays with temperature over twenty centigrade degrees, or not, etc.

With C1 and one of the criteria in C2 the analyst will form the different categories of holidays and will construct appropriate dummy variables for each category. For instance, with six days from C1 and two seasons, from November to April and from May to October, from C.2.a. one ends up with twelve categories of holidays.

Introducing the dummies for holidays the analyst would proceed to estimate the daily model and to check if for some holidays in the sample the corresponding residuals are in absolute value greater than three or four standard deviations. In the affirmative case these holidays will be treated with specific dummies for each of them. Holidays with specific effect use to be First of May and Easter Monday and Christmas Day. For specific holidays that can happen any day in the week like First of May the data usually would requize one dummy per day of the week. That, generally, means that the corresponding parameters would be estimated with only one degree of freedom. But if the differences from being a specific holiday in one day or another are big it is convenient to make such distinction for forecasting purposes. Sometimes it is enough to distinguished three or four days instead of six. Thus Cancelo and Espasa (1991) analyzing the electricity consumption in Spain consider four

different types of effect for First of May. They are, see table B.2 in the mentioned reference, -34% when First of May occurs on Monday, -35,4% when happens in Tuesday or Wednesday, -30,8% on Thursdays and Fridays and -5% in Saturdays. In this example the First of May effect on Mondays is not significantly different from Tuesday and Wednesdays but its dynamic effects -the effect on the day before and day after- are and it need a specific dummy.

Having detected specific holidays the analyst would use dummies for them and dummies for the holidays group in different categories. The next step should be to check if the different categories could be reduced testing if coefficients in the dummy variables are not significantly different.

Usually a holiday has a dynamic effect influencing the observation before and after the holiday. This should be taken into consideration when modelling these types of calendar effects by affecting the corresponding dummies by appropriate dynamic filters.

Finally if the population affected by a holiday could differ from one holiday to another the dummies should could be constructed taking into consideration the population involved in each case.

IX.2.- Vacation Periods.

The most relevant vacation periods are Easter, Christmas and Summer Holidays - month of August in several European countries-.

The effect of a vacation period will be modelled by the use of appropriate dummies, and to obtain them it would be necessary to test between different alternatives. It must be noted that these tests can not be carried over the whole set of residuals but only on the residuals corresponding to the vacation period. Otherwise with samples of two thousand residuals or more is impossible to discriminate between alternatives.

The alternatives for short vacation periods as Easter and Christmas could be defined according to the following criteria:

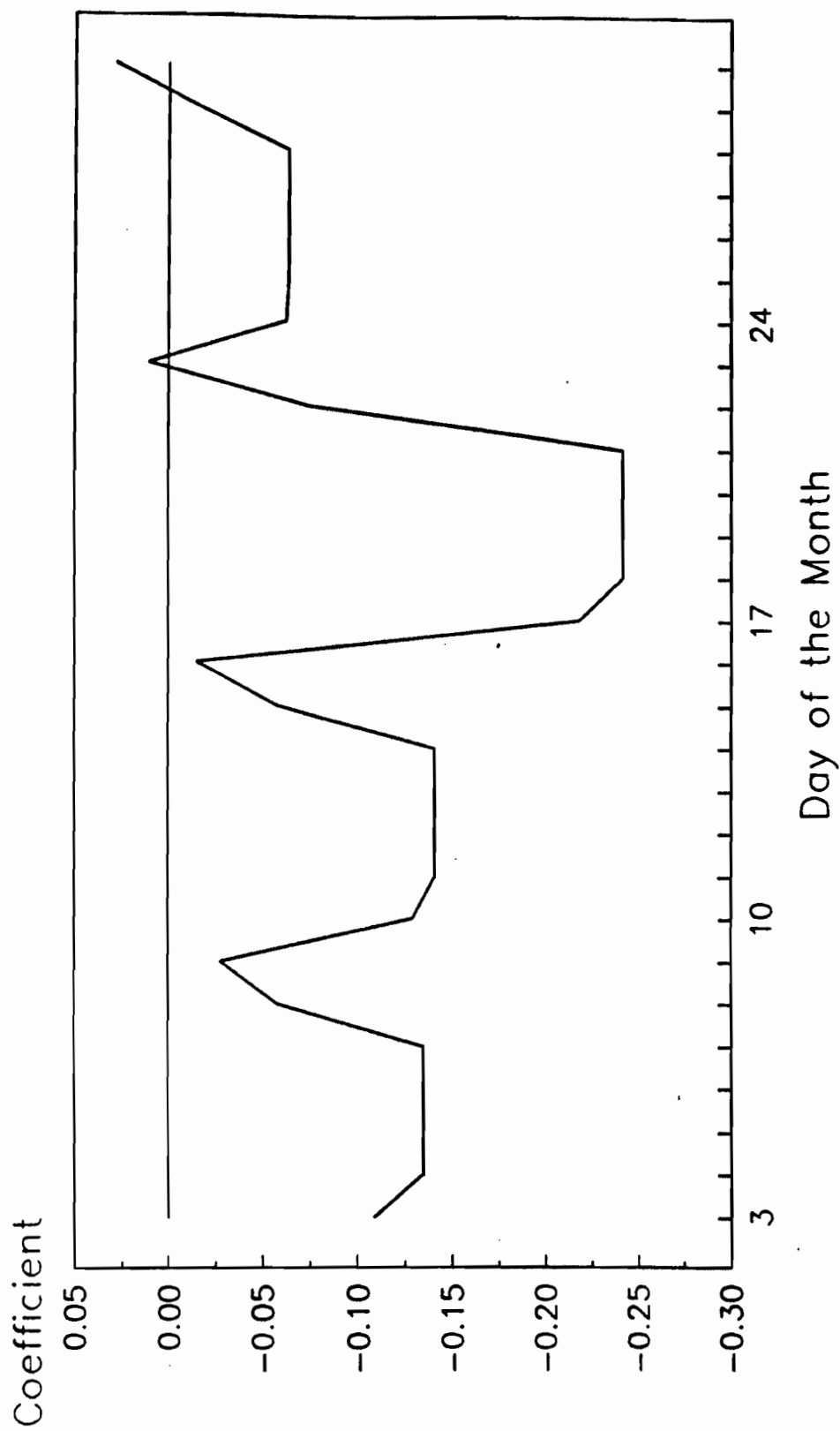
- (a) number of days involved in the vacation period;
- (b) a constant modification in the level of the daily data during the vacation period or a modification according to some functional or to some prior information;
- (c) specific effects of some holidays as Holy Friday, Easter Monday, Christmas Eve and Christmas Day, etc.

In longer vacation periods the main effects to capture are: (a) a change in the trend and (b) a change in the weekly cycle. Appropriate dummies for both things can be obtained using truncated step dummies all finishing at the end of the vacation period, and one of these steps starting at the beginning of the vacation period and the others each day in which a weekly seasonal modification is required. Figure 14 shows the reduction in electricity consumption in August estimated in Cancelo and Espasa (1991) using sixteen truncated step dummies starting each Monday, Tuesday, Saturday and Sunday of each of the four weeks of the vacation period. Imposing the restrictions that the steps of the second week have the same coefficients than the corresponding steps in the first week and that the steps in the third week have coefficients which doubled in value those of the first week, all the changes in trend and weekly seasonality can be captured with only eight coefficients.

IX.3.- Special Events.

The working calendar is altered by events as local or general elections, strikes, etc. They must be taken into consideration by specific dummies once the event has occurred. To forecast the impact of new event it is necessary to assume that it will be similar to the effect of a previous event or similar to the estimated effect of a holiday. For instance, the effect in the electricity consumption in Spain due to the general strike of Wednesday 13th December 1988 was quite well forecasted assuming that this strike would have the effect of a holiday in Wednesday.

FIGURE 14. — ESTIMATED EFFECT FOR AUGUST
(example) ϕ



One example of the importance of taking into consideration the calendar effects is in figure 15, taken from Cancelo and Espasa (1991b), which represents the stationary transformation of the electricity consumption (in logs) in Spain and the same transformation once the calendar effects have been removed from the data by appropriate dummies.

In cases the correction of data from the effect of special events is crucial. For instance, a strike in the banking sector could provoke that the daily figures reported as eligible habilities were wrong. In this case it could be important for a central bank to correct appropriately the reported figures in order to control the quantity of money in the system. Espasa (1979) shows how to estimate the effect of a strike of this type without estimating the whole dialy model.

X. Dynamic and nonlinear effects of meteorological variables.

Very often the effect of meteorological variables on daily data of economic activity is nonlinear and there can be different approaches to estimate it. Engle et al. (1986) propose a semiparametric method, Engle et al. (1992) use firts and second powers of the meteorological variable and Cancelo and Espasa (1991a) search for significant thresholds which define different segments in the range of variation of the corresponding meteorological variable. In this last case the authors approximate the nonlinear function in a piecewise linear way, allowing for different linear relationships for each one of the different significant segments in which the range of values of the explanatory variable can be divided.

Meteorological variables, specially temperature, can also have a dynamic effect on daily data. It means that a temperature of 12 centigrade degrees would have a different effect depending if the previous days have also registered 12 degrees, or a warmer or cooler temperature.

For modelling a nonlinear and dynamic effect the Cancelo and Espasa (1991a) approach has been proved quite useful. They specify a general dynamic response for each possible segment of temperatures, search for significant segments in order to approximate

FIGURE 15A.- DIFFERENCED ORIGINAL SERIES

(1983-1988)

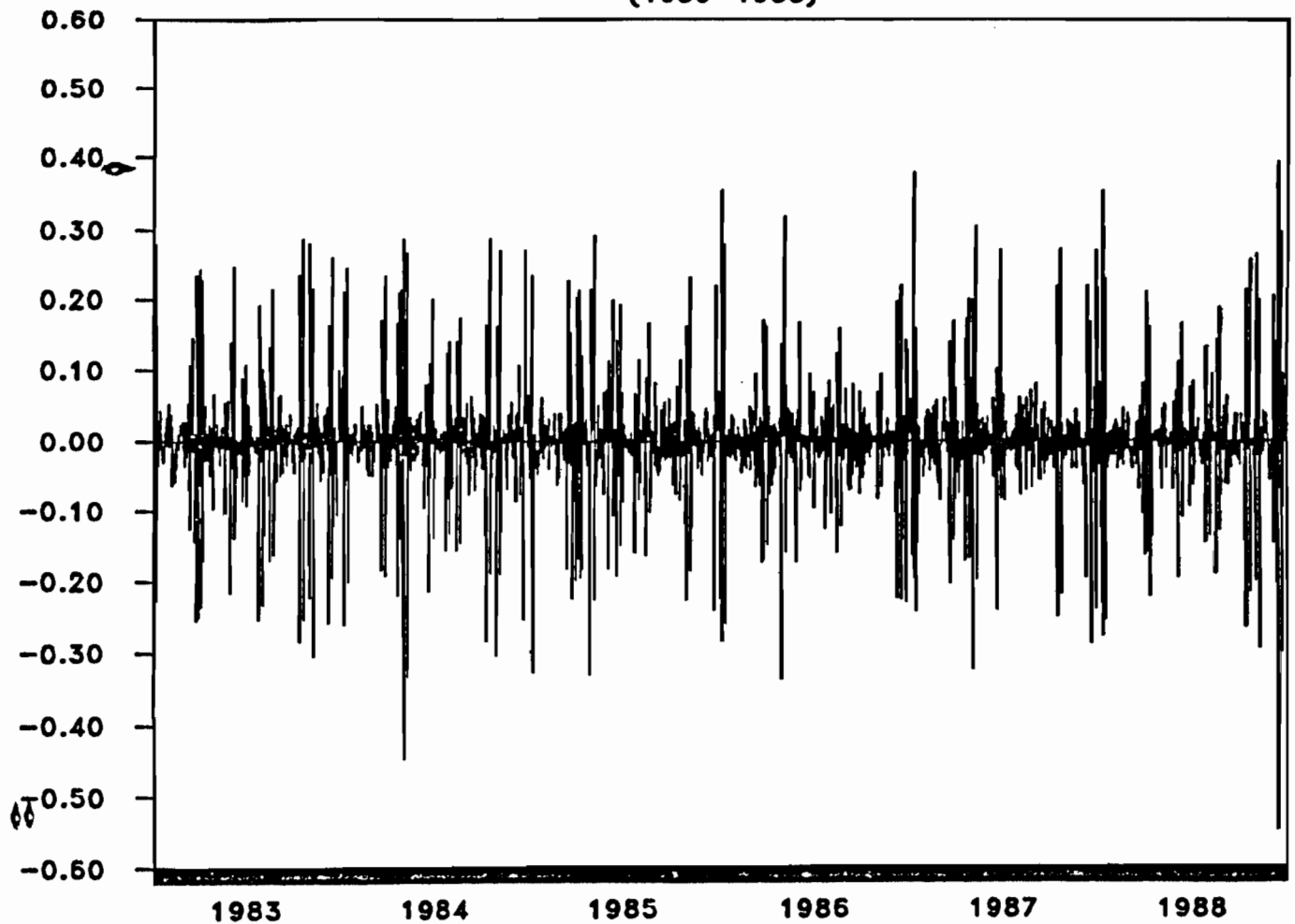
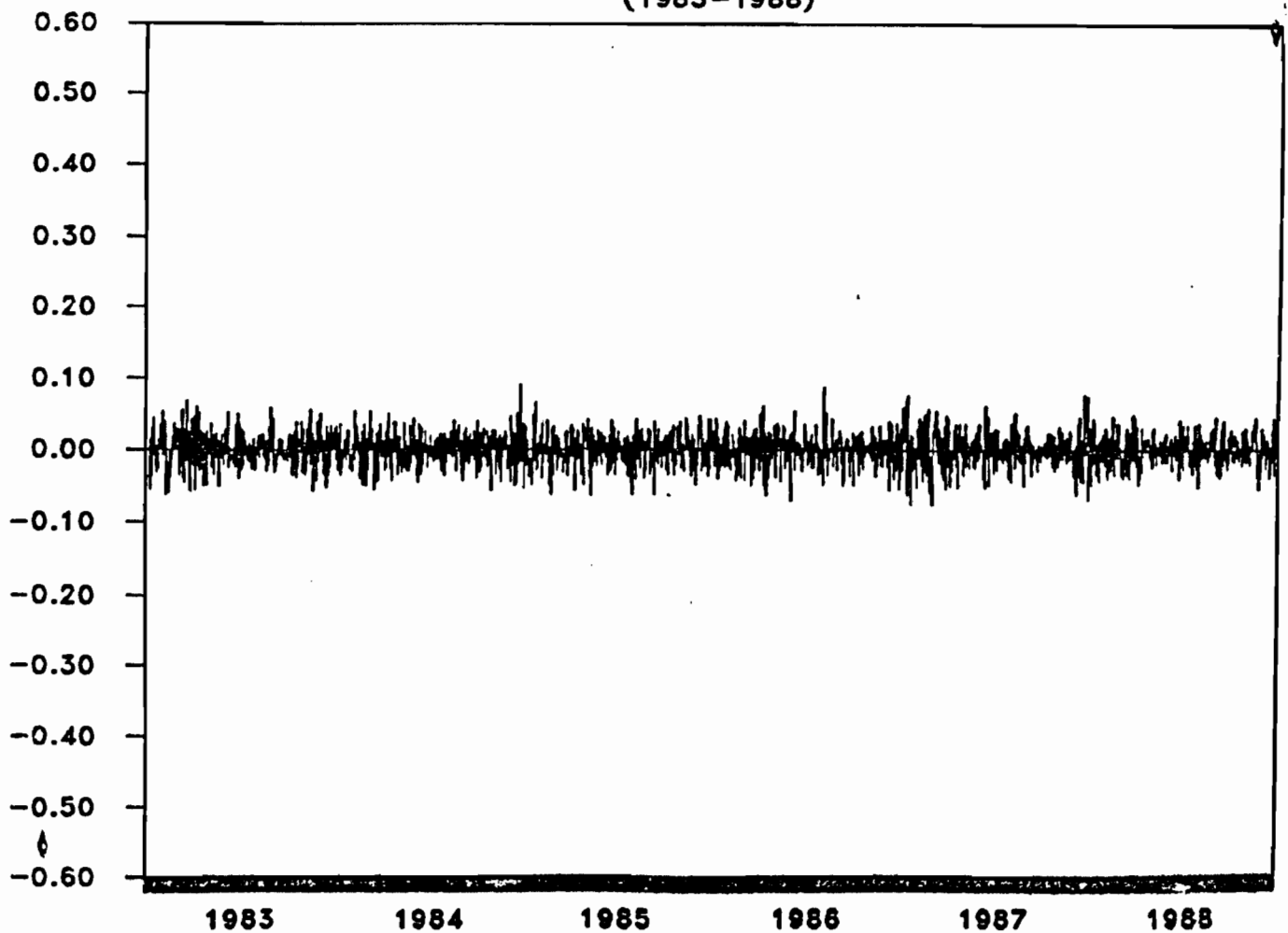


FIGURE 15B.- DIFFERENCED CORRECTED SERIES

(1983-1988)



the nonlinear function and in last step they test if the dynamic responses can be simplified. Thus the search procedure followed to specify the nonlinear dynamic relationship uses a general to specific approach for the dynamics and an specific to general one for nonlinearity.

In many cases this nonlinear dynamic response turn to be more complex because it changes with the season of the year and the type of the day, working days or holidays. Thus the elasticity on holidays uses to be greater than the elasticity for weekdays and the nonlinearity could be much more pronounced in winter then in the other seasons.

XI. The use of daily models for forecasting and control. Some practical examples from models in daily use from several years.

Cancelo and Espasa (1991b) present a model for a daily forecasting of the electricity consumption in Spain. The model has been estimated using 2859 observations and contains 161 parameters, 119 of them are required to model the calendar effects, 32 to model the effects of the different meteorological variables and 10 to specify an ARIMA model for the residuals. The standard deviation of the one day ahead forecast error is 1.2%. This model or previous versions of it is been used everyday for forecasting the electricity consumption in Spain since 1988. The model has predicted all the several historical maxima in electricity consumption that have occurred in Spain during this period, has produced accurate forecasts in days with unusual meteorological conditions and is very reliable in forecasting the consumption on holidays and vacation periods. The model, as it has been mentioned, was successfully used in forecasting the effect of the general strike which took place in Spain the Wednesday 13th December 1988 by assuming that the consumption this day would be like a holiday in Wednesday. Another important use of the model is for simulation. Thus the model can be used to simulate the extra consumption in a very cold winter or very hot summer or the save in electricity in a mild winter. Finally the model can be used to correct the daily data on electricity consumption from calendar effects and from the effect of abnormal meteorological conditions and employ the corrected data to construct weekly and monthly indicators of economic activity in Spain. This seems to be the quickest economic indicator that could be provided with almost a null cost.

For daily forecasting of notes and coin in circulation the Bank of Spain is using every days since 1986, see Cancelo and Espasa (1987), a model which in its last version contains 103 parameters, which was estimated using a sample of several years. The standard deviation of the one day ahead forecasting error is 0.14%. This data does not depend on meteorological variables but show end-and-beginning of month effects, different weekly cycles and a high sensitivity to holidays and Easter. The residuals are modelled by an ARIMA scheme with 12 parameters. The model is successfully used in forecasting several days ahead.

References

- Cancelo, J.R. and A. Espasa, 1987, "Un nuevo modelo diario para la predicción de la circulación fiduciaria", unpublished work, Banco de España, Servicio de Estudios.
- Cancelo, J.R. and A. Espasa, 1991a, "Threshold modelling of nonlinear dynamic relationships: an application to a daily series of economic activity", paper presented at the World Econometric Congress, September 1990, Barcelona. Working Paper 9105, Universidad Carlos III de Madrid.
- Cancelo, J.R. and A. Espasa, 1991b, "Forecasting daily demand for electricity with multiple-input nonlinear transfer function models: a case study", invited paper at the Eleventh International Symposium on Forecasting, New York. Working paper 9121, Universidad Carlos III de Madrid.
- Engle, R.F., C.W.J. Granger, J. Rice and H. Weiss, 1986, "Semiparametric estimates onf the relation between weather and electricity sales", Journal of the American Statistic Association, 81, 310-20.

Engle, R.F., C.W. Granger, R. Romanathan and F. Vahid-Araghi, 1992, "Probabilistic Methods in peak forecasting", final report on RP2919-04, Quantitative Economic Research Inc., 6095 Tamilynn St., San Diego, California.

Espasa, A., 1979, "Un modelo diario para la serie de depósitos en la banca: primeros resultados y estimación de los efectos de las huelgas de febrero de 1979", Working Paper 7905, Servicio de Estudios, Banco de España, Madrid.